UNIVERSITÉ DU QUÉBEC

THÈSE PRÉSENTÉE À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DU DOCTORAT EN PHILOSOPHIE

PAR
MICHEL PAQUETTE

LA DÉLIBÉRATION ET LES THÉORIES AXIOMATISÉES DE LA DÉCISION

JUILLET 2006

RÉSUMÉ

Cette thèse a pour objet l'explication de la délibération, c'est-à-dire, l'analyse des facteurs et des processus qui accompagnent une décision, dans le contexte des principales logiques axiomatisées de la décision qui sont discutées en philosophie. En examinant les axiomatisations de la logique de la décision qui se proposent d'expliquer le choix rationnel, il serait tentant de conclure que la théorie de la décision peut être formulée entièrement et adéquatement sans référence au processus de délibération de l'agent ou aux contraintes logiques qui s'exerce sur la délibération. Selon cette interprétation, le choix rationnel serait entièrement explicable dans le cadre de la théorie subjective des probabilités et de la théorie de l'utilité qui remonte à von Neumann et F. P. Ramsey. Cette interprétation de la logique de la décision a fait l'objet de nombreuses critiques durant la seconde moitié du XX^{e} siècle et en particulier durant les dernières décennies. Bien que la théorie classique impose le respect par ses applications, par sa simplicité et sa précision, la recherche de nouveaux modèles d'explication de la rationalité pratique est à l'ordre du jour. Cette thèse apporte une contribution à cette orientation de recherche. Nous soutenons que la logique de la décision, pour expliquer adéquatement le concept de rationalité pratique et surmonter les difficultés internes que signalent inlassablement ses critiques, doit mettre à profit les ressources des logiques philosophiques qui sont implicites dans son élaboration théorique et qui sont indispensables à l'expression de ses propres fondements. Il ne suffit pas de présupposer la logique classique pour

construire une logique de la décision. Il faut aussi intégrer explicitement la théorie des conditionnelles, les modalités temporelles et certaines contraintes épistémiques qui tiennent compte de la cinématique de la délibération et des capacités cognitives des agents. Nous étayons cette position logico-philosophique en proposant des arguments tirés principalement d'une étude détaillée des principales théories de la décision qui ont été proposées ou qui sont discutées en philosophie. Ainsi nous analysons les théories et commentons le modèle de la délibération qui est implicite dans la théorie de Ramsey (1926), de Carnap (1950), de Leonard Savage (1954/1972) ainsi que dans la logique de la décision de Bolker et Jeffrey (1965/1983). Par la suite, nous discutons de l'interprétation du paradoxe de Newcomb et nous en tirons des conséquences qui viennent renforcer et confirmer notre argumentation principale. Les théories causales de Gibbard et Harper (1978), D. Lewis (1981) et James M. Joyce (1999) sont passées en revue et comparées. Le dernier chapitre propose une synthèse des principaux résultats de notre recherche sur la délibération, présente une version de la logique du temps ramifié de Belnap et al. [2001] qui situe les agents les actions et les choix dans un monde indéterministe. De plus, ce chapitre propose un concept de décision qui peut s'exprimer dans un tel langage et qui est compatible avec la perspective philosophique développée dans cette thèse.

Je dédie ce travail à ma mère et à mon père,
Émilienne Doucet et Aurèle Paquette†.

REMERCIEMENTS

Je tiens à remercier les professeurs qui m'ont enseigné la logique durant mes études universitaires, Normand Lacharité (UQAM), Marc Venne (UQAM), Storrs McCall (McGill) et Anil Gupta (McGill). Je tiens aussi à remercier mon directeur de thèse, Daniel Vanderveken, qui a été un exemple et une source d'inspiration dans la poursuite de mes recherches. Enfin, je tiens à remercier mon épouse, Sylvie Lachize, pour ses encouragements et son soutien tout au long de la rédaction de cette thèse.

TABLE DES MATIÈRES

CHAPITRE I

INTRODUCTION

Les sciences n'essaient pas d'expliquer ; c'est tout juste si elles tentent d'interpréter ; elles font essentiellement des modèles. Par modèle, on entend une construction mathématique qui, à l'aide de certaines interprétations verbales, décrit les phénomènes observés.[1]

John von Neumann

1.1 La logique de la décision

Il y a un consensus pour reconnaître que l'explication logico-philosophique de concepts fondamentaux a connu au moins deux succès retentissants au XX^{e} siècle. Le premier de ces succès est l'explication du concept de nombre cardinal par Frege et Russell au début du siècle ; le second est l'explication du concept de vérité par Tarski. L'explication du choix rationnel en logique de la décision est, de l'avis de plusieurs, un troisième accomplissement du même ordre. Comme le note Davidson la théorie de la décision est, sur le plan méthodologique, comparable à celle de Tarski et elle devrait être évaluée de la même façon[2]. En se proposant de la défendre contre les critiques qui voudraient faire table rase, Brian Skyrms rappelle que la théorie de la décision développée pas Ramsey, De Finetti et Savage est une « partie importante de notre héritage intellectuel »[3]. Dans un passage dithyrambique de *Reason and*

Human Affairs, H. Simon parle des théories formelles de la rationalité comme de « joyaux parmi les accomplissements intellectuels de notre temps »[4]. Si la théorie de la décision peut trouver sa place à côté des contributions de Frege et Russell ou de celle de Tarski, elle mérite effectivement notre attention. C'est dans cet esprit que nous avons abordé la logique de la décision et nous proposons cette perspective comme un point de départ et une idée directrice pour décrire la méthodologie de la présente thèse. En effet, l'explication formelle d'un concept apporte avec elle des critères et des objectifs qu'il est bon de rappeler[5].

Dans le même esprit que la théorie de la vérité de Tarski, la logique de la décision ne cherche pas à utiliser des termes familiers, tels les termes « décision » et « choix rationnel » pour décrire des concepts nouveaux, mais elle tente plutôt de saisir le sens effectif de ces vieilles notions. Le problème principal est de donner une définition satisfaisante de la notion de choix rationnel, autrement dit une définition qui soit matériellement adéquate et formellement correcte. Comme l'a vu Carnap, il est possible que ce processus de clarification conduise à remplacer une notion vague par un seul concept plus exact. En logique de la décision, il semble plutôt que le concept initial soit remplacé par un ensemble de concepts plus précis. Pour fournir ce type d'explication, il faut construire un modèle au sens particulier que prend cette expression dans la citation de von Neumann que nous avons mise en épigraphe. En logique de la décision, ce modèle prend la forme d'une construction logique et mathématique qui donne la valeur d'une option en combinant d'autres constructions comme des préférences et des fonctions de probabilité. La logique de la décision fournit également un ensemble de contraintes qui donnent une structure logique à la hiérarchie des préférences.

Nous pouvons préciser ce qui précède en formulant une définition générale de la logique de la décision au sens de la présente thèse. Elle propose un cadre théorique pour expliquer le choix rationnel et utilise une combinaison de constructions théoriques de façon à prescrire comment des agents individuels devraient choisir lorsqu'ils sont dans l'incertitude et qu'ils possèdent une connaissance incomplète des facteurs qui déterminent les résultats de leurs actions possibles[6].

Examinons les composantes de cette définition dans l'ordre. La logique de la décision se réclame d'un concept de choix rationnel qui possède une longue tradition. Au XVII[e] siècle à l'époque où Pascal découvrait le calcul des probabilités, on retrouve ce concept clairement exprimé dans la *Logique de Port-Royal*:

> [...] pour juger de ce que l'on doit faire pour obtenir un bien, ou pour éviter un mal, il ne faut pas seulement considérer le bien et le mal en soi, mais aussi la probabilité qu'il arrive ou n'arrive pas; et regarder géométriquement la proportion que toutes les choses ont ensemble.[7]

Le choix rationnel est donc celui qui combine la désirabilité d'un bien avec la probabilité de l'obtenir. On voit que la construction théorique nécessaire pour expliquer ce concept demande de combiner les croyances au sujet de ce qui est désirable, ce qu'on appelle une théorie des préférences et un jugement de probabilité, ce que Ramsey appellera une théorie de la croyance partielle. Cette théorie est entièrement tournée vers l'appréciation de l'avenir car, comme l'a remarqué Aristote, on ne peut délibérer qu'au sujet du futur. Bien que nous n'ayons pas encore expliqué et justifié ce parti pris, nous devons souligner que la définition propose de considérer la logique de la

décision comme une théorie normative qui indique ce qu'un agent[8] devrait faire. C'est pourquoi la définition utilise le verbe « prescrire ». Enfin, nous disons que l'agent à propos duquel se pose la question de la rationalité d'un choix est un être incertain. En effet, pour un agent omniscient ayant la capacité d'anticiper le futur, le problème de la rationalité instrumentale, c'est-à-dire le problème de choisir le moyen le plus efficace pour arriver à ses fins ne se poserait pas. Il saurait toujours d'avance ce qu'il doit faire.

Le concept de rationalité est un des concepts centraux du schème conceptuel de la philosophie du langage et de la cognition. La philosophie du langage, la philosophie de l'esprit et l'épistémologie utilisent ou présupposent un tel concept. Il est appelé à jouer un rôle dans l'élaboration de la plupart des constructions théoriques qui se proposent d'interpréter ou d'expliquer le comportement humain. Considérons par exemple l'étude du langage qui a une importance spéciale en philosophie analytique. Le philosophe Tyler Burge a déjà exprimé cette omniprésence de la raison dans le comportement humain en observant que lorsqu'on fait la rencontre d'une personne entièrement inconnue, simplement parce qu'elle parle une langue naturelle et par ce seul fait, on sait déjà qu'elle est rationnelle.

Il est manifeste que les locuteurs compétents et efficaces d'une langue naturelle sont rationnels. Ils utilisent leur rationalité pour planifier leurs actes de discours, les choisir et interpréter les actes de référence et de prédication qu'effectuent leurs interlocuteurs. Qui plus est, leur rationalité leur permet de suivre le fil d'une conversation, d'anticiper la suite et de sélectionner des actes de discours potentiellement pertinents qui vont engendrer des contributions réussies aux échanges conversationnels. Dans la typologie des discours proposée par Daniel Vanderveken, la notion de délibération est prise comme

un des types de base pour catégoriser plusieurs buts discursifs possibles. Il existe plusieurs sortes de discours qui concernent les délibérations individuelles ou celles des groupes. Le but délibératif concerne les actions possibles que les locuteurs et les allocutaires s'engagent à accomplir dans le monde. Ce but discursif qui corrrespond à la délibération a une direction d'ajustement qui va des choses vers les mots[9]. L'application du concept de rationalité à la pragmatique du discours, à l'analyse de l'argumentation et la logique du dialogue compte parmi les retombées plus lointaines de la présente recherche. La théorie des jeux est un autre domaine de recherche qui possède des liens étroits avec la logique de la décision et qui a une grande importance pour la philosophie. Dans le passé, la théorie des jeux s'est développée conjointement avec la logique de la décision et elle a utilisé les mêmes constructions[10]. Ainsi, on doit considérer la théorie des jeux comme un domaine connexe, voire inséparable, de la logique de la décision. Les remarques qui précèdent indiquent sommairement les domaines où les progrès en logique de la décision peuvent avoir des retombées.

1.2 Enjeux méthodologiques

Dans la présente thèse, il ne sera question que de choix individuels. Il s'agit d'un choix méthodologique qui découle d'un principe de division du travail et ce choix ne doit pas être interprété autrement. Ainsi, nous ne croyons pas que le choix collectif soit réductible au choix individuel, que la délibération collective pose moins de difficultés théoriques que la délibération individuelle ou que les questions qu'elle soulève soient moins importantes. Dans le même ordre d'idée, pour prévenir tout malentendu, nous n'allons pas

tenter de décrire le comportement des personnes ni chercher à répondre à la question piège de savoir si les humains sont rationnels. Cette question, tout compte fait, n'est pas si intéressante qu'elle paraît l'être. Car, comme l'a remarqué Davidson, on ne décrit pas le comportement d'un agent, on l'interprète[11]. Ainsi, on peut imaginer qu'en étant suffisamment charitable, il serait possible de réconcilier tout comportement d'un agent avec une norme soigneusement choisie et formulée pour pouvoir admettre une interprétation favorable. Lorsque nous disons que la logique de la décision est une théorie normative, nous entendons par là qu'elle l'est de la même façon que la logique des propositions. En conséquence nous acceptons aussi que la possibilité que certaines de ses règles puissent faire l'objet de discussions philosophiques[12].

On reproche parfois à la théorie de la décision de ne s'appliquer véritablement qu'à des agents idéalisés qui seraient des surhommes de sagesse. En réalité, c'est tout le contraire. Comme nous l'avons signalé, le problème du choix rationnel concerne des agents qui sont incertains et qui ne sont pas omniscients. Dans la théorie de la décision de David Lewis, un détail technique fait en sorte que la théorie ne pourrait pas s'appliquer à un agent parfaitement rationnel. En réaction à cette difficulté, il remarque qu'il est difficile d'imaginer qu'un tel agent pourrait se retrouver dans la position de délibérer car il possède une compréhension parfaite des relations entre ses actions possibles et leurs effets[13].

Un très grand nombre de publications ont critiqué la logique de la décision en lui opposant des études empiriques qui montrent manifestement que les sujets étudiés ne respectent pas les préceptes de la théorie, l'exemple le mieux connu d'un tel précepte étant sans doute la transitivité des préférences[14]. On désigne ce débat comme le contraste entre l'interprétation

descriptive et l'interprétation normative de la théorie de la décision. Cette question a fait rage durant de nombreuses années, mais la place accordée à ces débats dans les revues philosophiques tend à diminuer.

David Lewis a bien exprimé le sens qu'il faut accorder aux modèles de la logique de la décision de façon à éviter l'obstacle que pose l'adéquation descriptive. Après réflexion, il nous semble que sa façon de représenter les relations entre les constructions mathématiques de la logique de la décision et la réalité mentale des agents dispose efficacement des reproches qui concernent le réalisme cognitif de la théorie. La remarque qui suit concerne la formulation de la théorie causale de la décision mais elle représente de façon perspicace une interprétation principalement normative de la logique de la décision.

> J'ai idéalisé la situation et simplifié outrageusement de trois façons, mais je crois que les complications que j'ai escamotées ne font aucune différence à propos de la question de savoir pourquoi et comment la théorie de la décision doit être causale. D'abord, il semble très invraisemblable qu'une personne réelle puisse mémoriser et traiter quoi que ce soit qui s'approche des fonctions C et V que nous considérons. On est contraint de faire avec des versions sommaires. *Mais il est plausible qu'une personne qui aurait véritablement ces fonctions pour le guider ne se conduirait pas de façon bien différente, sauf pour ses prouesses de logique, de mathématiques et, de façon générale, pour ses connaissances a priori.*[15]

À notre avis cette interprétation indique clairement le critère d'adéquation qui est pressenti par une interprétation normative de la logique de la décision. Il existe un raisonnement imparable dû à Van McGee, raisonnement qui s'appuie sur deux théorèmes normalement démontrés, à l'effet que même un

agent qui aurait les capacités d'une machine de Turing ne pourrait se conformer à la norme qui demande de maximiser l'utilité espérée[16]. Nous reconnaissons entièrement le bien-fondé de cette argumentation de McGee sauf au dernier pas lorsqu'il pose comme une vérité universelle que «je dois» (*ought*) implique « je peux » (*can*). Les humains doivent être rationnels, ils doivent s'efforcer de l'être, mais il n'est pas invraisemblable que ce soit au-dessus de leurs capacités[17]. Dans les chapitres II à IV de la présente thèse, nous montrerons précisément comment Ramsey, Carnap, Savage et Jeffrey ont abordé la relation entre leurs théories et les capacités cognitives des agents.

D'autres critiques de la théorie de la décision ont posé des problèmes qui se donnent comme des contre-exemples à la théorie de la décision. Cependant il arrive fréquemment que ces contre-exemples s'écroulent lorsque l'on reformule le problème. En réalité, comme nous allons le constater à quelques reprises, les problèmes de décision sont sensibles à leur formulation et en recadrant une situation, un choix apparemment sous-optimal est validé par la théorie. Le philosophe John Searle raconte que lorsqu'il a été initié à la théorie de la décision, il s'est trouvé étonné d'une conséquence de la théorie apparemment contraire à une intuition ferme :

> Il semble que ce soit une conséquence stricte des axiomes que si j'accorde une certaine valeur à ma vie et que j'accorde une valeur à une pièce de 25 ¢ (une pièce de 25 ¢ n'a pas beaucoup de valeur mais elle en a suffisamment pour être ramassée sur le trottoir, par exemple), il doit exister une cote pour laquelle je serais disposer à parier ma vie contre une pièce de 25 ¢.[18]

Searle ajoute que cette caractéristique lui a suffi pour rejeter la théorie et se convaincre qu'il y avait un grave problème avec la théorie de la décision. Il

déplore que certains des plus grands théoriciens de la décision furent incapables de lui expliquer cette anomalie. Nous ne partageons pas l'avis de Searle. Nous nous rangeons plutôt du côté de Davidson, du côté de la plupart des théoriciens de la décision, pour répondre que personne n'a avancé une idée claire de ce qui pourrait montrer que la théorie de la décision est fausse. N'oublions pas qu'il s'agit d'une théorie normative[19].

L'énoncé de ce problème invite le lecteur à imaginer une situation dramatique, hors de l'ordinaire, peut-être la fameuse scène du film *The Deer Hunter* où le héros est un prisonnier forcé de jouer à la roulette russe durant la guerre du Vietnam. Qui, et dans quel contexte imaginable, parierait sa vie contre une pièce de 25 ¢ ? En réalité, la situation qui permet de résoudre l'énigme du point de vue de la théorie de la décision n'a rien de dramatique et elle n'exige aucune prouesse de l'imagination. En fait, on peut la trouver *dans* l'anecdote que raconte Searle. Si j'accepte de me pencher pour ramasser une pièce de 25 ¢ sur le trottoir, je suis disposé à parier ma vie contre cette somme. Comment ? C'est que, par le simple fait de me pencher, j'augmente ma vulnérabilité à une attaque par laquelle un agresseur pourrait me tuer. C'est peu probable dites-vous ? Je suis entièrement d'accord avec vous. Et c'est en cela que réside la clef de l'énigme. Les êtres humains ne sont pas très doués pour estimer des probabilités très faibles. Lorsqu'elles sont quasi négligeables, elles sont régulièrement négligées. C'est pourquoi nous avons du mal à admettre l'idée que nous risquons notre vie à des degrés divers en traversant la rue, en conduisant une voiture ou en allant à la banque. Somme toute, ce qu'exige la théorie de la décision est d'accepter le slogan qui veut qu'une action délibérée puisse toujours se décrire *d'une certaine façon* comme un pari[20]. Il n'y a donc aucune anomalie dans l'exemple de Searle. La morale de

cette histoire concerne aussi la sensibilité de tout problème de décision à la manière dont il est formulé ; si je change la formulation, il arrive souvent que je change le problème. Comme le dit Joyce

> Le choix est véritablement un processus qui se fait en deux étapes dans lesquelles l'agent commence par affiner ses idées sur la situation de décision dans laquelle elle se trouve en examinant soigneusement ses options et l'état du monde jusqu'à ce qu'elle soit fixée sur le « vrai » problème de décision (...) toute analyse complète doit avoir quelque chose à dire à son sujet.[21]

Cette observation de Joyce nous conduit naturellement à reconnaître l'importance de la délibération dans le choix rationnel. Comme nous le verrons, si cette place a été négligée par les théoriciens de la décision, c'est en partie parce que la méthode axiomatique invite à minimiser le rôle des interprétations verbales qui accompagnent les axiomes. Ce sera notre tâche de les découvrir et de les exposer. Il se peut aussi que ce soit pour éviter la complexité des diverses logiques philosophiques qui sont nécessaires pour rendre explicite les modalités de la délibération. Nous utiliserons fréquemment l'expression « infrastructure logique » pour désigner l'ensemble de toutes ces logiques (logique du temps ramifié, logique de l'action, logique épistémique, théorie des conditionnelles) qui selon nous, doivent permettre ultimement de résoudre les difficultés réelles ou appréhendées des théories que nous discutons. En fin de parcours, nous esquisserons les constituants élémentaires de cette infrastructure logique. Pour l'explication de la délibération, nous proposerons que la logique du temps ramifié est un élément essentiel de

l'appareil logique requis parce qu'il permet de situer les agents et leurs actions dans un monde indéterministe.

1.3 L'argument principal

La thèse que nous défendons concerne l'importance de la délibération dans la construction d'une logique de la décision. Cette thèse qui s'inscrit dans les retombées des enseignements du problème de Newcomb peut sembler plausible d'un point de vue intuitif. Cependant, un partisan irréductible de telle ou telle version des théories classiques pourrait nous répondre par une objection qui doit être prise en sérieuse considération. En effet, l'examen des axiomatiques de la décision semble montrer que la logique de la décision tente par tous les moyens de valider un argument simple qui ruinerait notre approche. Selon cet argument, une fois que les préférences et les probabilités subjectives sont déterminées, tout ce qui importe pour fixer la valeur d'une option est déterminé. Cet argument, s'il s'avérait probant, irait manifestement dans le sens opposé de notre hypothèse de recherche. Nous appellerons l'énoncé catégorique de cet argument *l'argument principal.*

Pour les défenseurs d'une théorie non-causale de la décision, si vous êtes un agent parfaitement rationnel, vos choix sont entièrement guidés par vos désirs et vos croyances ainsi que votre capacité d'introspection[22]. Ainsi, si vous êtes entièrement rationnels, rien ne peut influencer vos choix si ce n'est en exerçant une influence sur vos croyances et vos désirs. Pour définir le choix rationnel, il suffirait de tenir compte de ces deux paramètres. Les autres aspects de la délibération ne seraient pas pertinents et n'influenceraient pas la valeur d'une option.

Comme nous le verrons aux chapitres IV et V, l'étude des problèmes qui s'apparentent au problème de Newcomb ne se laisse pas circonscrire dans le cadre étroit que voudrait imposer l'argument principal. À notre avis, cette étude rend l'argument principal insoutenable. Nous discuterons aussi de nombreuses difficultés qui interpellent l'infrastructure logique de la théorie de la décision et qui constituent autant d'objections à cette position réductrice.

1.4 Présentation des chapitres

Il existe un grand nombre d'axiomatisations de la logique de la décision et P. Fishburn expose sans doute les plus importantes dans un article qui propose une analyse comparée et un bilan en date de 1981[23]. Nous en mentionnerons d'autres plus récentes au chapitre V. Parce que nous nous intéressons à l'explication philosophique du concept de choix rationnel, nous prêtons une grande attention aux interprétations verbales qui accompagnent les constructions mathématiques. Ces explications nous livrent souvent le sens et la portée de ce qui est accompli ainsi que quelques indications sur les rapports qui relient ces constructions à des agents réels ou possibles[24]. Pour cette raison, les théories que nous allons examiner dans les prochains chapitres ont presque toutes été proposées par des philosophes. On peut d'ailleurs les situer de façon successive selon l'ordre chronologique[25]. Ainsi nous analyserons au chapitre II la théorie de F. P. Ramsey (1926) et pour plusieurs raisons nous accordons une grande importance à sa contribution. Nous étudierons l'évolution de la pensée de Carnap sur le choix rationnel et en particulier, sa conception des probabilités. La conception subjective des probabilités qui est utilisée en théorie de la décision est encore méconnue à

l'extérieur des départements de philosophie et d'économie et nous en avons étudié les fondements au cours de notre recherche. Comme nous le verrons, la position pluraliste qui est commune à Ramsey et Carnap permet de contourner ces débats qui concernent des enjeux secondaires relativement à nos objectifs. Selon ces auteurs, la reconnaissance du bien-fondé de la conception subjective n'implique pas que nous devions nier l'acceptabilité de la conception fréquentiste. Ainsi, la validité de l'interprétation des probabilités en termes de fréquence qui est utilisée dans plusieurs sciences empiriques peut être pleinement reconnue dans le contexte de son application.

Le chapitre IV porte sur la théorie de Leonard Savage (1954/1972) ainsi que la logique de la décision de Richard Jeffrey (1965/1983) axiomatisée grâce à une contribution mathématique due à Ethan Bolker. Ces théories constituent le point de référence de toutes les discussions à propos de la logique de la décision en philosophie[26]. En apparence, ces théories sont similaires, mais lorsqu'on y regarde de près, des différences importantes apparaissent tant au niveau de l'analyse du choix que des principes qu'elles valident. Leur infrastructure logique est plus détaillée que celles de toutes les théories antérieures.

Au chapitre V, nous analyserons en détail la problématique de l'interprétation du paradoxe de Newcomb et nous en tirons des conséquences qui viennent renforcer et confirmer notre argumentation principale à propos de l'importance de la délibération. Même si nous croyons que le problème de Newcomb est incontournable et que sa solution nous conduit du côté des défenseurs de la théorie causale, nous soutiendrons la viabilité de la position de ceux qui rejettent le problème, position qualifiée de *no-boxer*. Les théories causales de Gibbard et Harper (1978), D. Lewis (1981) et James M. Joyce

(1999) sont passées en revue et comparées. Lorsqu'on veut représenter la délibération qui conduit à une décision, on arrive à un conditionnel. La forme logique de ce conditionnel est l'objet d'un débat ; est-ce plutôt le mode indicatif qui s'exprimerait par « si je fais ceci il se produit cela » ou la forme subjonctive « si je faisais ceci, il se produirait cela ». À la suite de Joyce, nous montrerons que cette question trace une ligne de partage entre les théories évidentielles et les théories causales.

Le VI[e] et dernier chapitre propose une synthèse des principaux résultats de notre recherche sur la délibération. Il présente une version de la logique du temps ramifié de Belnap et al. [2001] qui situe les agents les actions et les choix dans un monde indéterministe. De plus ce chapitre propose un concept de décision qui peut s'exprimer dans un tel langage et qui est compatible avec la perspective philosophique développée dans cette thèse.

1.5 Conventions et typographie

Nous utilisons les expressions formées d'un nom propre suivi d'une année entre crochet pour désigner une entrée dans la bibliographie. Dans chaque chapitre, la notation utilisée est presque toujours identique à celle de la théorie qui est à l'étude, de façon délibérée. Le but est de faciliter la comparaison avec les textes originaux. Par conséquent, il n'y a pas de cohérence notationnelle d'un chapitre à l'autre, mais seulement un effort de cohérence à l'intérieur de chaque chapitre. Nous utilisons souvent des variantes alphabétiques différentes des formules originales pour les constantes propositionnelles ou autres. Lorsqu'il nous semblait qu'aucune confusion n'était possible, nous n'avons pas utilisé les guillemets pour distinguer l'usage

d'une expression de sa mention. À quelques endroits, nous utilisons les guillemets anglais pour exprimer la prudence. Selon l'usage, les notes de fin de document ne sont jamais nécessaires à l'intelligence du propos. Elles comportent néanmoins des commentaires et les références sont presque toujours données à titre de justifications secondaires. L'usage de néologismes est toujours accompagné d'une note qui signale son inévitabilité. Lorsqu'elles peuvent contribuer à la compréhension, les expressions originales anglaises sont mentionnées en italique entre parenthèses. La typographie est ajustée aux indications du cahier des *Normes de Présentation des Travaux de Recherche du Décanat des Études avancées et de la Recherche* (UQTR) complétées par les indications du *Lexique des règles typographiques en usage à l'imprimerie nationale.*

CHAPITRE II

LA THEORIE DE F. P. RAMSEY

2.1 Importance historique de la théorie de Ramsey

Rédigé en 1926, l'essai intitulé « Truth and probability » est un des textes fondateurs de la conception subjective des probabilités au XXe siècle. On y trouve une des premières formulations de la définition du concept de probabilité en termes de degré de croyance. Bien sûr, les questions historiques d'antériorité sont délicates et il faut se garder de poser en ces matières des jugements catégoriques. Ainsi, l'idée de mesurer le degré de conviction par la disposition à parier est assez ancienne et on la retrouve clairement énoncée chez Kant[1]. Mais avant Ramsey, il semble que seul Émile Borel[2], dans un compte rendu de l'ouvrage de J.M. Keynes [1921], aurait eu l'idée d'utiliser des paris pour mesurer les degrés de croyance dans le but de définir le concept de probabilité[3]. Par ailleurs, à peine deux années plus tard, Bruno de Finetti va concevoir séparément le même concept de probabilité subjective et il publiera ce résultat neuf ans plus tard dans de Finetti [1937]. Par la suite, von Neumann et Morgenstern [1944] ont redécouvert indépendamment les idées principales de Ramsey. Enfin, on retrouvera des conceptions comparables de la croyance et de la probabilité axiomatisée dans Savage [1954] ainsi que dans Suppes et Davidson [1956] qui se réfèrent explicitement à l'essai de Ramsey et qui ont contribué à le faire connaître. L'importance de l'essai de Ramsey est

également mise en évidence par la discussion critique dont il fait l'objet dans Jeffrey [1965][4].

La conception subjective de la probabilité est plus souvent associée au nom de Bruno de Finetti qu'à celui de Ramsey et l'importance historique de l'essai de Ramsey n'a été reconnue que tardivement. D'ailleurs, cet essai est relativement peu cité dans la littérature sur la théorie de la décision ou à l'extérieur du domaine de la philosophie. Pourtant, nous croyons qu'on ne saurait trop souligner l'importance et les mérites de l'essai de Ramsey ; la profondeur des thèses qui y sont avancées et la valeur des argumentations qui y sont développées en font un essai logico-philosophique exemplaire et de première importance pour notre recherche. On a d'ailleurs qualifié ce texte de « remarquable » ou « d'extraordinaire »[5]. Il faut savoir que Ramsey avait une compréhension parfaite des *Principia Mathematica* de A. N. Whitehead et B. Russell ainsi que du *Tractatus logico-philosophicus* de L. Wittgenstein dont il a réalisé la première traduction anglaise. C'est sur cet arrière-plan théorique qu'il faut comprendre son effort pour élargir le domaine de la logique au sens le plus général du terme. Son objectif est de construire une logique de la croyance partielle qui repose sur des fondements logico-mathématiques exacts et aussi explicites que possible. C'est pourquoi nous le considérons comme le fondateur du champ d'investigation dans lequel s'inscrit notre propre recherche. On sait que Ramsey est mort très jeune en 1930 à l'âge de 26 ans, et qu'il travaillait à la conception d'un ouvrage entièrement consacré aux concepts de vérité, de croyance et de probabilité. Ainsi, l'article « Truth and probability » ne représente pas un traitement définitif des concepts qu'il discute. Ramsey écrira plus tard qu'il n'est pas entièrement satisfait de son explication du concept de probabilité, principalement parce qu'il le trouve trop

psychologique[6]. Il croit cependant avoir jeté les bases d'une nouvelle façon d'aborder ce concept:

> Je n'ai pas construit en détail la logique mathématique de ceci [la logique des degrés de croyance] car ce serait, à mon avis, comme construire jusqu'à sept décimales un résultat qui n'est valide qu'à la deuxième. Ma logique ne peut pas être considérée comme exprimant plus qu'une indication de la façon dont ça pourrait fonctionner.[7]

Nous verrons que c'est bien ce qu'il aura réalisé, autant par son explication des probabilités que par son développement du concept de degré de croyance interprété en termes de paris. Son analyse de ces concepts constitue clairement un point de départ pour étudier la conception subjective des probabilités et la logique de la décision. On peut aussi soutenir qu'elle fournit des fondements logico-philosophiques solides pour une théorie de la décision et que ces fondements sont encore pertinents pour les débats actuels[8]. Dans les sections qui suivent, nous présentons les idées principales de la théorie de Ramsey en essayant de dégager des choix méthodologiques et des principes logico-philosophiques qui seront utiles par la suite. Principalement centrés sur l'essai Ramsey [1926], mon exposé et mes commentaires s'appuient sur quelques sources secondaires qui seront citées au passage dont les principales sont Sahlin [1990] et Sobel [1998].

2.2 Conceptions rivales de la probabilité

Considérons d'abord les choses du point de vue actuel. De nos jours, on oppose le plus souvent deux conceptions des probabilités, l'interprétation

objective et l'interprétation subjective. On dit des probabilités qu'elles sont *objectives* si on les interprète comme des fréquences observées que nous révèle l'expérience et l'on dit qu'elles sont *subjectives* en tant qu'elles sont les modalités de l'expression de croyances incertaines. Ramsey croyait à la légitimité de deux concepts différents de probabilité: le concept de probabilité comme fréquence et le concept de probabilité comme degré de croyance qu'il va analyser et défendre. Pour Ramsey, l'interprétation des probabilités en termes de fréquences a ses mérites. D'une part, elle est conforme à l'usage intuitif du terme « probable » qui en fait une sorte de proportion. D'autre part, elle offre une structure d'interprétation simple pour les lois du calcul de probabilité. De plus, il semble que l'interprétation en termes de fréquences soit l'interprétation visée par l'usage généralisé des probabilités de la science moderne. De nos jours, suivant de Finetti [1937], un courant subjectiviste s'est développé qui refuse entièrement toute interprétation objective des probabilités. Quoi qu'il en soit de la possibilité d'éliminer le concept de probabilité en tant que fréquences observées, le langage ordinaire et une longue tradition nous autorisent également à considérer la théorie des probabilités dans une autre optique, c'est-à-dire en tant que logique de la croyance partielle. C'est ce concept qui intéresse Ramsey. Il faut souligner que Ramsey n'a pas inventé l'idée d'interpréter les probabilités subjectivement. À la suite de Hacking [1975], on peut dire que la double nature des probabilités est aussi vieille que l'idée de probabilité elle-même[9].

À l'époque où il écrivait son essai sur la vérité et la probabilité, deux conceptions principales des probabilités étaient discutées : la conception fréquentielle[10], la mieux connue, et la conception logique proposée par Keynes [1921]. Il faut se garder de confondre la conception logique des

probabilités associée aux noms de Keynes et Carnap et la conception subjective. Pour cette raison, nous devons d'abord éclaircir cette différence entre la conception logique et la conception fréquentielle. Nous n'aurons plus à reparler de la conception fréquentielle par la suite. C'est la critique de la conception logique de Keynes qui permet à Ramsey de poser son approche. Pour Keynes, les probabilités sont des relations qui existent objectivement et elles sont définies pour toutes les paires de propositions. En ce sens, elles se comparent aux relations logiques de compatibilité ou de déductibilité par exemple. Plus précisément, la conception élaborée par Keynes de l'interprétation logique des relations de probabilité pose qu'entre deux propositions prises respectivement comme prémisse et conclusion, il y a une et une seule relation constante α telle que si nous avons une croyance totale en la prémisse, nous devrions, si nous sommes rationnels, avoir une croyance de degré α dans la conclusion.

Pour illustrer la différence entre l'interprétation fréquentielle et l'interprétation logique, considérons une pièce de monnaie pour laquelle nous supposons qu'une fois lancée en l'air, elle retombera éventuellement sur l'une ou l'autre face, sans jamais rester en équilibre sur la tranche, de façon à ce que la probabilité qu'une telle pièce une fois lancée retombe sur pile est de 0,5[11]. Selon la formule consacrée de l'interprétation fréquentielle, la valeur fractionnaire 1/2 représente la tendance du rapport entre le nombre de cas favorables et le nombre de cas possibles. Dans un traité bien connu d'introduction à la théorie des probabilités dû à J. E. Freund, on trouve cette définition de l'interprétation fréquentielle qui exprime essentiellement la même idée :

> [...] **l'interprétation en termes de fréquences**, selon laquelle la probabilité d'un événement (d'un résultat) est interprétée comme la proportion du nombre de fois où des événements similaires vont se produire sur une longue période.[12]

Si je lance la pièce un grand nombre de fois, la proportion de succès va tendre à se rapprocher de plus en plus de 0,5. Ce concept de probabilité est bien celui qui est utilisé par la science empirique selon l'interprétation habituelle et c'est pour lui que la statistique a développé ses techniques.

En reprenant la formule de Keynes, une interprétation logique de cet exemple poserait qu'entre les propositions 2.1 et 2.2,

(2.1) La pièce de monnaie lancée peut retomber du côté pile ou du côté face et pas autrement.

(2.2) La pièce retombera du côté pile.

il existe une relation logique qui fait en sorte qu'en acceptant le premier énoncé, (2.1), comme prémisse, si on est rationnel, on doit accorder une valeur de crédibilité de 0,5 au second, (2.2). La probabilité est ici considérée comme une propriété logique entièrement expliquée par le rapport entre la conclusion et la prémisse. Notons au passage que Keynes utilise le concept de fonction propositionnelle abondamment mais il ne clarifie pas ce qu'il entend par proposition. En lisant Keynes [1921], on peut penser qu'une proposition est une fonction propositionnelle n-aire saturée ou un énoncé d'une langue naturelle qui exprime une telle forme logique.

2.3 Critique de Keynes et nouveaux fondements

Comme on l'a vu plus haut, avant d'introduire sa propre conception des probabilités, Ramsey discute et rejette cette conception logique de la probabilité de Keynes. L'examen et la critique de la conception de Keynes sont importants parce qu'ils permettent d'aborder directement la différence de nature entre la logique classique, c'est-à-dire la logique déductive telle que formalisée par A. N. Whitehead et Russell dans les *Principia Mathematica* et la logique de l'inférence probable[13]. En cherchant à axiomatiser le calcul des probabilités, il semble que Keynes cherchait à faire pour le raisonnement probable ce que Russell, à la suite de Frege, avait fait pour le raisonnement déductif. D'entrée de jeu, il peut sembler attrayant de chercher à construire une explication de la probabilité dans le cadre du logicisme, c'est-à-dire de chercher à expliquer les probabilités uniquement à partir des propriétés des propositions et des relations qu'elles ont entre elles. C'est l'échec de ce programme qui conduira aussi bien Ramsey que de Finetti et plus tard, comme on le verra au chapitre III, Carnap, à la conception subjective.

Nous serons en mesure de mieux caractériser l'objectif et l'approche de Keynes après avoir introduit quelques précisions terminologiques. Le terme « induction » est défini de façon différente selon les auteurs et l'inférence inductive est parfois assimilée à l'inférence probable. L'usage de l'expression « probabilité inductive » est aussi de nature à introduire une confusion.

Considérons d'abord le concept d'induction. Évitons de donner un sens restrictif à ce terme et reprenons plutôt l'explication qu'en donne Brian Skyrms dans le *Cambridge Dictionary of Philosophy* :

Par *induction*, on entend, (1) au sens étroit, une inférence qui conclut à une généralisation à partir de ses instances ; (2) au sens large, toute *inférence ampliative*, c'est-à-dire toute inférence dont la thèse affirmée en conclusion va

plus loin que la thèse affirmée conjointement par les prémisses. La logique inductive, au sens le plus large, est la théorie de l'évaluation des inférences ampliatives.

Voyons maintenant les définitions proposées par Keynes. Pour lui, l'induction et l'analogie relèvent de « la Probabilité », qu'il définit comme « cette partie de la logique qui étudie les raisonnements qui sont rationnels mais non-concluants »[14].

Keynes distingue deux sortes d'inférences inductives : *l'induction universelle*, qui conclut à une loi et dont la conclusion est une proposition universelle, par exemple, « Tous les cygnes sont des oiseaux » et la *corrélation inductive* dont la conclusion utilise une quantification moindre, telle « La plupart des cygnes sont blancs »[15]. Dans les deux cas nous dit-il, un raisonnement inductif affirme, non pas qu'un certain état de choses existe, mais plutôt que sur la base d'un certain nombre de faits observés (*evidence*) il y a une probabilité en sa faveur.

Dans les premières pages de son traité, Keynes explique clairement qu'il prend comme point de départ l'inférence probable et qu'il veut en tirer une clarification de la nature des probabilités. Son objectif plus lointain est de s'attaquer au problème de l'induction au moyen d'une analyse de l'inférence probable.

Keynes, Ramsey ou Skyrms conviennent tous qu'un raisonnement n'est pas entièrement formé de la liste des énoncés qui forment ses prémisses et sa conclusion. On doit aussi considérer la nature de la relation entre les prémisses et la conclusion, relation par laquelle les prémisses donnent un poids relatif à la conclusion.

Comme nous venons de le voir, l'induction est une inférence non-démonstrative. C'est cependant une erreur courante que de croire que toutes les inférences non-démonstratives sont inductives[16]. Le raisonnement par analogie est aussi une inférence non-démonstrative. Qui plus est, on remarque que ce ne sont pas toutes les inférences probables qui sont ampliatives. De « La plupart des oiseaux volent », on peut conclure que « Quelques oiseaux volent ». Ici, la conclusion en dit moins que la prémisse. Le concept d'inférence probable est donc plus général que le concept d'induction. C'est pourquoi, comme Ramsey, nous préférons utiliser l'expression « inférence probable » plutôt que de proposer une définition inhabituelle de l'expression « induction » qui inclurait les inférences non-ampliatives[17].

Il y a deux autres raisons qui nous incite à éviter le terme d'induction. La première raison est que nous considérons, à la suite de Hume, l'induction comme une habitude de l'esprit. Nous utilisons tous des inférences inductives ; elles peuvent créer en nous des convictions, mais la justification des inductions ne peut pas être donnée en les réduisant à des inférences déductives ou dans le cadre limité de la logique formelle. On peut aujourd'hui considérer comme acquis que cet ancien programme de la logique inductive n'est pas viable. Ceci ne signifie pas pour autant qu'il faille renoncer à l'espoir de bien caractériser la force des inférences probables. C'est le domaine de la logique probabiliste (probability logic) qui a été bien développé dans les travaux d'Hugues Leblanc et de Charles Morgan par exemple[18].

Notre seconde raison additionnelle pour éviter la problématique de l'induction est liée à l'énigme de l'induction. On parle parfois d'induction et de logique inductive en relation avec le *problème traditionnel de l'induction*,

la tentative pour établir la validité de ce que Keynes appelle l'induction universelle, mieux connue sous le nom d'induction naturelle[19]. Un tel usage est validée par la définition du *Cambridge Dictionary of Philosophy* que venons de citer. Malheureusement, nous ne voyons aucun mérite aux tentatives pour valider la tendance naturelle à généraliser. Lorsque des méthodes inductives sont encadrées par des règles précises dans le cadre d'une méthode de recherche en sciences, ces méthodes peuvent être jugées fiables, mais on a débordé le cadre de la logique et on se trouve d'emblée sur le terrain de la méthodologie et de l'épistémologie.

On peut contraster l'explication que je viens de donner de l'induction et de l'inférence probable avec le système de définitions proposé par Brian Skyrms dans son ouvrage introductif, *Choice and Chance*. Ayant posé que la logique étudie le lien entre les prémisses et la conclusion d'un raisonnement, Skyrms introduit le concept de « force inductive » d'un raisonnement pour les raisonnements non-démonstratifs. Un raisonnement non-démonstratif est « inductivement fort » s'il est improbable que la conclusion soit fausse si les prémisses sont vraies. Dans le cas de l'inférence non-démonstrative, puisque la valeur du raisonnement (c.-à-d., la nature du lien entre les prémisses et la conclusion) est une question de probabilité, on peut définir la « probabilité inductive » comme une mesure de la force inductive d'un raisonnement. Cependant, comme le fait justement remarquer Skyrms, il n'est pas possible de clarifier directement ce concept de probabilité inductive[20]. Pour obtenir une telle clarification, il faut construire une *logique inductive* et idéalement une *logique inductive scientifique* qui encode nos pratiques scientifiques d'inférence inductive.

L'approche de Skyrms associe l'inférence probable à l'induction par le biais du concept de force inductive. Elle pose le problème de ce que nous appelons l'inférence probable de façon purement logique plutôt qu'épistémologique ou méthodologique mais elle reporte le problème d'une définition *précise* de la probabilité inductive. Ainsi, Skyrms conclut que la solution complète de ce problème doit être trouvée au-delà de la logique proprement dite comme il l'indique dans la définition suivante.

> Convenons d'appeler *logique inductive* une définition précise de la probabilité inductive, associée à une méthode pour déterminer la probabilité inductive de raisonnements ainsi que des règles pour construire des raisonnements inductivement forts.[21]

Cette méthode d'évaluation et ces règles de construction doivent être de nature méthodologique et épistémologique. À mon avis, cette façon de voir les choses rappelle l'idée de *méthode inductive* chez Carnap puisqu'on ne pose pas qu'il y a une méthode unique qui soit valide et parce qu'on laisse la nature exacte de ces méthodes dans l'arrière-plan. Il est très clair que Skyrms ne fait pas de confusion entre la logique et l'épistémologie ou entre l'inférence non-démonstrative et le problème classique de l'induction. Il indique lui-même que les distinctions qu'il pose sont approximatives (*shrewd*). Il me semble que cette façon de définir la logique inductive et le rôle des probabilités n'est pas très utile dans le présent contexte. En premier lieu, l'opposition faite par Skyrms entre probabilité inductive et probabilité épistémique ne coïncide pas avec ce qui est en question dans le débat entre Ramsey et Keynes. Telle que définie, la probabilité épistémique est plus large que la probabilité subjective (ou personnelle) puisqu'elle permet une conception de la probabilité où la

base est la totalité du savoir humain ou la totalité du savoir d'un être omniscient. Nous ne croyons pas que Skyrms endosse une telle position mais elle est concevable si on s'en tient à l'opposition probabilité inductive/probabilité épistémique. Aussi, la probabilité inductive au sens de Skyrms diffère de la probabilité logique de Keynes car il n'y a aucune indication que la probabilité inductive doive être objective et a priori. En s'appuyant sur ses autres écrits, on peut affirmer sans crainte que Skyrms ne souscrit pas à une interprétation objectiviste comme celle de Keynes. En second lieu, l'expression « force inductive » d'un raisonnement peut laisser croire qu'il s'agit d'analyser l'induction, ce qui n'est pas le cas comme le révèle une lecture attentive. En troisième lieu, l'expression « probabilité inductive » suggère qu'on a affaire à une sorte de probabilité, c'est-à-dire, à une explication du concept de probabilité. Or, ce n'est pas ça non plus. Skyrms présente le problème de l'interprétation des probabilités séparément en présentant les diverses possibilités d'un point de vue impartial. Avec ces clarifications en tête, on peut retourner à la critique de Keynes par Ramsey.

La première difficulté soulevée par Ramsey[22] à l'endroit de la théorie de Keynes concerne cette relation objective de probabilité qui serait de nature logique et qui existerait entre un ensemble de propositions désignées comme les prémisses d'un raisonnement et une conclusion[23]. Il n'est pas exagéré d'affirmer, comme le fait Ramsey, que dans la perspective très particulière de Keynes, les relations de probabilité existent entre toutes les propositions, de façon indépendante des esprits *et* des inférences que font réellement les gens.

Keynes croit que les degrés de croyance et les relations de probabilité ne peuvent pas toujours s'exprimer par des nombres. S'il en est ainsi, « les probabilités ne peuvent pas être identifiées aux degrés de croyance qui les

justifient »[24] car une mesure du degré de croyance par un « psychogalvanomètre » (*sic*) ne serait pas acceptable comme mesure indirecte de relations de probabilité qui sont censées être objectives. Au fond, il apparaît clairement que le point central est la tension apparente qui existe entre « l'objectivité » de la relation recherchée et son identification à un degré de croyance. C'est cette tension que Ramsey va révéler et mettre en évidence par ses critiques. Il se déclare incapable de percevoir les relations de probabilité dont parle Keynes et soupçonne que personne d'autre ne les perçoit tant il est difficile de se mettre d'accord sur la relation de probabilité qui existerait entre deux propositions données. Nous avons donc, d'une part, une théorie mathématique qui est fort développée et qui nous dit comment additionner et multiplier les probabilités, et d'autre part, le contenu réel extrêmement mince et fuyant qui en serait le pendant réel selon Keynes. C'est, écrit Ramsey, comme si nous connaissions les lois de la géométrie, mais que personne ne pouvait dire si un objet donné est carré ou circulaire[25].

Cette critique conduit à douter de l'existence de relations de probabilité conçues comme des relations logiques, uniques et objectives par suite de la difficulté de les percevoir clairement et de s'entendre sur leur degré. Si on rejette l'hypothèse de Keynes, ceci ne signifie pas pour autant qu'il faut renoncer à expliquer l'inférence probable. Car si Keynes a tort et qu'il n'existe pas une relation unique de probabilité qui mesurerait exactement le rapport logique entre les prémisses et la conclusion d'un raisonnement non-démonstratif, cette relation n'est pas pour autant inexplicable. Ramsey indique comment aborder la question de l'inférence probable en expliquant intuitivement le sens d'une conditionnelle qui joue un rôle central dans sa conception :

> Il me semble que si on prend deux propositions « *a* est rouge », « *b* est rouge », on ne peut discerner plus de quatre relations logiques simples ; ce sont l'identité de forme, l'identité du prédicat, la diversité des sujets et l'indépendance logique de leurs contenus respectifs. Si quelqu'un devait me demander quelle probabilité l'une donne à l'autre, je ne chercherais pas à répondre en contemplant les propositions et en essayant de discerner une relation logique qui existe entre elles ; je chercherais plutôt à imaginer que la première proposition est tout ce que je sais et je me demanderais quelle confiance je devrais avoir en l'autre.[26]

En lisant la première partie de cette citation, on constate que les relations logiques simples, lorsqu'elles existent, sont clairement perceptibles. La seconde partie de cette citation donne un principe fondamental de la conception subjective. Les jugements de probabilité sont toujours relatifs à une base de connaissance, qui forme la sphère de croyance de l'agent[27]. De plus, si je peux estimer la confiance que j'ai dans le second énoncé, je peux aussi admettre que je fasse erreur dans cette évaluation et j'en arrive à m'inquiéter de l'évaluation que ferait quelqu'un de plus sage que moi. Nous aurons l'occasion de revenir sur cette suggestion qui nous force à considérer la possibilité que c'est l'évaluation d'un sujet idéal qui serait le degré de probabilité de la proposition. Lorsque nous sommes tous d'accord, par exemple, pour estimer que la probabilité que la pièce de monnaie lancée retombe du côté pile est de 0,5, personne ne semble en mesure de préciser exactement la classe des énoncés empiriques qui constitue le fondement de cette évaluation. On est forcé de constater que cette observation est simultanément une objection à la conception de Keynes et un argument fondamental pour la conception subjective des probabilités.

Pour mettre en évidence l'importance théorique de la conditionnelle de Ramsey qui semble correspondre à une intuition robuste, nous devons souligner avec insistance sa pertinence du point de vue de la logique philosophique. Considérons d'abord la formulation qu'il en donne dans un autre texte :

> ***Conditionnelle de Ramsey :*** Si deux personnes débattent de la question « Si p, est-ce que q adviendra ? » et qu'elles sont toutes les deux dans le doute à propos de p, elles ajoutent p hypothétiquement à leur bagage de connaissance et argumentent sur cette base à propos de q ; de telle sorte qu'en un sens, "Si p, q" et "si p, $\neg q$" sont contradictoires. On peut dire qu'elles fixent leur degré de croyance en q étant donné p [...].[28]

Ce que nous appelons « conditionnelle de Ramsey » peut être considéré comme la formulation du sens d'une conditionnelle, une sorte de « si...alors... ». Les choix et les paris que nous discuterons plus loin, aux sections 6 et 7, doivent s'interpréter selon ce principe. Or, cette conditionnelle ne correspond manifestement pas à l'implication matérielle de la logique classique[29]. La recherche d'une explication exacte, c'est-à-dire d'une formalisation, de ce type d'implication a connu des développements subséquents dans le cadre d'une logique épistémique interprétée en termes de mondes possibles ou dans le cadre de l'étude des connecteurs d'une logique dont l'interprétation visée est fournie par des fonctions de probabilité[30]. Nous reviendrons plus loin sur les acquis et les difficultés de cette branche importante de la logique philosophique. Il est clair que la conditionnelle de Ramsey correspond à une idée importante pour l'élaboration d'une logique de la décision. Il faut prendre garde au fait que la difficulté de capturer le sens de

cette relation dans une logique formelle constitue un défi à relever plutôt qu'une difficulté conceptuelle d'ordre philosophique qui soit spécifique à l'approche de Ramsey. Nous avons tenté de montrer la différence entre ce programme de recherche et le problème central de la logique inductive. L'ambition du programme classique de la logique inductive était de valider le principe de l'induction naturelle, c'est-à-dire, le passage d'un certain nombre d'énoncés d'observation, qui sont pris comme prémisses, à une proposition universelle qui serait la conclusion. Il nous semble que la recherche d'une telle logique inductive relève de plein droit de l'épistémologie au sens traditionnel d'une *philosophie de la connaissance* et au sens contemporain de *philosophie des sciences*. En ce sens, il s'agit d'un immense problème et il est clair que la logique inductive engendre un vaste programme de recherche. Ce programme, dont la viabilité n'est pas clairement établie par des considérations d'ordre logique, est différent de la problématique de la présente recherche. Par conséquent, il n'est pas souhaitable de lier notre entreprise à la logique inductive et à la posture empiriste qui lui donne son sens et sa portée. L'intention est d'analyser l'inférence probable comme élément constitutif d'une logique de la décision et du choix rationnel. Pour cette raison, nous ne voulons pas associer notre démarche à une posture épistémologique particulière. Un des objectifs de Keynes était de valider le principe de l'induction naturelle. Comme nous venons de l'indiquer, nous voudrions éviter de faire basculer la logique de la décision dans ce programme de recherche sans nier qu'il en est voisin[31].

Contre la théorie de Keynes, Ramsey poursuit en analysant ce qui semble une incohérence dans sa présentation. D'une part, Keynes est disposé à admettre que nous ne connaissons pas (et que nous ne pouvons pas connaître)

la relation de probabilité qui existe entre la conclusion d'une inférence et ses prémisses[32]. Mais, dit Ramsey, si nous ne la connaissons pas, nous ne pouvons pas en tirer le degré de croyance qu'il est légitime de placer dans la conclusion. Il y a donc une disparité entre les relations de probabilité logiques inconnaissables de Keynes et les degrés de croyance qu'elles devraient légitimer. Cette disparité suggère de chercher une autre conception de la probabilité qui se passerait entièrement des relations de probabilité indéfinissables pour ne conserver que les degrés de croyance rationnelle des agents. C'est cette voie que nous allons maintenant examiner.

2.4 La croyance partielle et sa mesure

Ayant rejeté la conception de Keynes, Ramsey s'attaque au concept de degré de croyance, concept qui deviendra central dans la recherche de nouveaux fondements pour l'explication de l'inférence probable. Il s'agit d'un concept de nature psychologique et cette nature doit être reconnue de plein droit pour qu'il soit possible d'en faire une analyse adéquate. La difficulté qui surgit immédiatement est celle de la détermination d'une échelle graduée pour comparer les degrés de croyance et le problème de les quantifier. Si le concept de degré de croyance doit recevoir un sens opératoire et jouer un rôle dans l'explication des probabilités, ce concept doit être clarifié et la notion de « degré » doit recevoir une interprétation qui implique la possibilité de comparer et d'associer des nombres. Aussi, comme le remarque Ramsey

> Il ne suffit pas de mesurer la probabilité ; pour ajuster correctement notre croyance à la probabilité, nous devons aussi pouvoir mesurer notre croyance.[33]

Or, il est clair qu'une quantification est possible pour certains degrés de croyance. C'est le cas, par exemple, pour la croyance totale en une proposition qui se voit associer la valeur 1, et la croyance totale en sa contradictoire qui reçoit la valeur 0. La valeur 1/2 est associée à une croyance égale en la proposition et en sa contradictoire. Ceci peut s'exprimer par les formules suivantes où l'on pose que pour un agent quelconque :

1^{er} cas : $Cr^{\circ}(p) = 1$ (croyance totale en p)

2^{e} cas : $Cr^{\circ}(\neg p) = 0$ (croyance totale en l'impossibilité de p)

3^{e} cas : $Cr^{\circ}(p) = Cr^{\circ}(\neg p) = 1/2$ (incertitude complète à l'égard de p)

Il est moins facile d'interpréter la signification d'une croyance dont la valeur serait 2/3. Comme toutes les autres croyances partielles, elle correspond à l'incertitude. Sur le plan philosophique, il va s'avérer impossible d'écarter entièrement la difficulté liée à l'interprétation de ce genre de quantification et l'on constate dans la littérature qu'elle revient sans cesse au premier rang des critiques à propos du concept de croyance partielle et dans les discussions de la conception subjective des probabilités. On trouve une discussion des mérites de cette critique et une défense comparable à celle de Ramsey chez la plupart des tenants de la conception subjective. De plus, cette critique peut être formulée à propos d'autres concepts dont la quantification détaillée ne va pas de soi ou pour lesquels la détermination d'une échelle comparative est problématique[34]. Ramsey et ceux qui le suivront vont soutenir que la détermination complète des valeurs de l'échelle numérique pour le degré de croyance et le concept corrélatif de probabilité ne sont pas requis pour valider

ces concepts. Tout cela est bien consonant avec ce que nous présentions dans le chapitre précédent sous l'appellation « théorie qualitative de la décision ». Cette défense est donc d'une importance capitale et l'argumentation mérite d'être étudiée soigneusement.

Keynes croyait que certains degrés de croyance sont mesurables et que d'autres ne le sont pas. Ramsey rejette d'emblée cette idée. Pour lui, le fait que certaines croyances sont plus aisément mesurables que d'autres indique que le procédé par lequel on mesure le degré de croyance est changeant et que ce procédé de mesure engendre des valeurs différentes selon la façon de mesurer. Il exploite une analogie avec le concept d'intervalle de temps dans la physique d'Einstein pour illustrer que le concept de degré de croyance n'a pas de signification précise tant que n'est pas précisé le procédé de mesure. Comme pour le concept d'intervalle dans le contexte de la physique classique, on peut supposer que les diverses façons de mesurer le degré de croyance vont coïncider mais ceci n'est qu'une approximation qui simplifie l'analyse. La façon de procéder à cette mesure et le choix d'une unité de mesure sont les premiers problèmes à considérer soigneusement.

Cette tâche se divise en deux. Dans un premier temps, un système adéquat de mesure du degré de croyance doit assigner une position précise sur une échelle de grandeur. En idéalisant un peu, on peut accepter l'idée de croyances de degrés égaux et ces croyances doivent recevoir une même position dans l'échelle de grandeur. On pourra ainsi construire une suite ordonnée. Dans un second temps, il faut associer des nombres à ces degrés de façon intelligible et perspicace.

La solution de Ramsey repose sur une conception pragmatique de la croyance. Il ne s'agit pas d'aborder le degré de croyance comme une intensité

émotionnelle ressentie qui accompagne la croyance. Il s'agit plutôt d'évaluer le degré de la croyance par la disposition à agir qu'elle induit, c'est-à-dire, par son efficacité causale[35]. Pour soutenir cette approche, Ramsey répond aux arguments de Russell qui a discuté et rejeté la définition pragmatique de la croyance[36]. Bien sûr, ce ne sont pas toutes nos croyances qui sont à l'origine d'actions. Mais la conception pragmatique retenue ici ne fait qu'affirmer que la croyance effective *conduirait* à l'action dans les circonstances appropriées. Il n'est pas non plus requis de supposer que les croyances ne se distinguent les unes des autres que par leur efficacité causale. Mais comme l'efficacité causale est la face observable de la croyance, c'est la face qui importe pour lui assigner une interprétation quantifiée[37]. On aura l'occasion de revenir sur l'évaluation critique de la conception pragmatique de la croyance dans les chapitres suivants de la thèse en examinant la construction d'autres modèles du choix rationnel. Comme Ramsey, nous ne croyons pas que cette conception soit nécessairement réductrice. En conséquence, la question principale que nous aurons à examiner sera celle de savoir si la conception pragmatique de la croyance peut constituer un sous-ensemble d'une explication plus substantielle.

Pour obtenir une théorie de la croyance partielle, Ramsey va s'inspirer de la méthode qui consiste à proposer à l'agent un pari, il va en généraliser le principe et la rendre plus exacte. Plus exacte, car il faut contourner la difficulté associée à la diminution de l'utilité marginale de l'argent et au fait qu'un agent peut être plus ou moins disposé à parier selon qu'il a plus ou moins le tempérament d'un joueur. Il faut aussi la généraliser à toutes les croyances et à toutes les valeurs de l'échelle des probabilités et non seulement considérer les propositions et les valeurs qui s'interprètent plus intuitivement

comme les termes d'un pari possible. Pour y arriver, Ramsey adopte une théorie psychologique minimaliste et idéalisée :

> [...] la théorie qui veut que nous agissions de la façon qui va le plus vraisemblablement nous permettre de réaliser nos désirs, de telle sorte que les actions d'une personne sont complètement déterminées par ses désirs et ses opinions.[38]

Les difficultés apparentes de cette théorie psychologique ne sont pas niées, mais il est montré qu'elles ne sont pas dirimantes. Par exemple, cette théorie n'interdit pas que nous agissions parfois dans le dessein d'obtenir un bien pour en faire profiter autrui et elle n'exclut pas que nous agissions pour des motifs inconscients ou des opinions inconscientes. C'est en utilisant une analogie avec la mécanique newtonienne dans sa relation à la théorie de la relativité que Ramsey revendique la légitimité d'utiliser une théorie dont la vérité n'est qu'approximative et il note qu'il serait souhaitable que le noyau de vérité de ce système artificiel de psychologie (*sic*) soit préservé par une théorie psychologique plus élaborée.

En introduisant une mesure du degré de croyance, il devient possible de préciser cette théorie psychologique. Appelons « biens », ce que les personnes désirent comme finalités ultimes et convenons que les biens sont quantifiables (numériquement mesurables) et additifs. Ce faisant, on postule que si une personne préfère une heure de natation à une heure de lecture, il va préférer faire deux heures de natation plutôt que de faire une période d'une heure de natation suivie d'une période d'une heure de lecture. Cet exemple est intentionnellement contraire à l'intuition et sert à souligner qu'il s'agit d'une idéalisation. Il est clair qu'après une heure de natation, une heure de lecture

peut sembler plus attrayante qu'une deuxième heure de natation. Ramsey explique cette observation en disant que la lecture et la natation ne sont pas des biens ultimes (*ultimate goods*). Ramsey ne clarifie pas le concept de bien ultime et il ne nous donne aucun exemple. Le contexte dans lequel il utilise l'expression indique que c'est une propriété essentielle d'un bien ultime que la valeur associée au fait de disposer d'une $x^{ième}$ quantité de ce bien ne sera pas moindre que la valeur d'en posséder une seule unité. Un bien ultime ne doit pas être affecté par le phénomène de la diminution de son utilité marginale lorsque sa quantité augmente. Comme l'ont d'abord fait remarquer Gabriel Cramer et Daniel Bernoulli, l'argent n'a pas une utilité marginale constante. Il en va de même pour la lecture ou la natation. De plus, on remarque que ces activités ne sont pas des biens dont la désirabilité est intrinsèque. Ainsi, on valorise la lecture pour l'information qu'elle procure, le divertissement ou la réflexion qu'elle permet. On apprécie la natation pour ses effets bénéfiques sur le plan physique ou sur le plan moral. Ces biens ne sont pas ultimes car nous ne les valorisons pas pour eux-mêmes. C'est bien le sens qu'il faut associer au qualificatif « ultime ». Un bien ultime serait un bien que l'on désire pour lui-même, qui possède une valeur intrinsèque, c'est-à-dire dont la valeur est celle d'une finalité plutôt que celle d'un moyen. Peut-on trouver un exemple de ce que serait un bien ultime ? On pourrait penser qu'une heure de temps libre est un bien que l'on désire pour sa valeur intrinsèque mais ce serait oublier que l'addition d'heures de temps libre engendre l'ennui. Le seul exemple qui semble résister à ce type d'objection est le bonheur. Le bonheur en tant qu'état général de bien-être est sans doute le meilleur exemple d'un bien qu'on valorise pour lui-même et qui ne perd pas de sa valeur alors que sa quantité augmente. Il est clair que l'existence d'un exemple indiscutable de bien ultime

n'est pas une objection substantielle dans la construction de Ramsey. Tout comme le mètre pourrait jouer le rôle d'unité de mesure même si aucun objet réel ne mesurait exactement un mètre, le bien ultime peut jouer le rôle d'*équivalent général* pour l'échelle d'utilité sans qu'il soit nécessaire que des exemples de tels biens existent.

Pour expliquer le degré de croyance en termes de pari, on procède en deux étapes. Dans un premier temps, on pose par hypothèse que l'agent a une croyance associée à toutes les propositions. On pourrait aussi dire, ce qui semble logiquement équivalent, qu'il ne doute de rien. Cette hypothèse ne résiste pas à un examen critique et elle ne se justifie que par la simplification qu'elle permet. En effet, une tête bien éduquée devrait avoir pris en considération au moins quelques propositions logiquement indécidables. De plus, il y a des propositions extrêmement longues ou complexes qu'aucun agent ne peut se représenter mentalement et qui, pour cette raison, ne peuvent faire l'objet d'une attitude épistémique comme la croyance. Ici encore, on ne peut pas élaborer une objection substantielle à la théorie de Ramsey sur la base de ces remarques. Ramsey ne fait qu'utiliser une hypothèse dont la raison d'être est de simplifier la construction de sa théorie.

La seconde étape pour obtenir les degrés de croyance partielle est l'utilisation du concept d'espérance mathématique[39]. Ce concept de la théorie des probabilités provient de l'analyse des jeux de hasard. La personne qui possède un billet de loterie pouvant lui permettre de gagner un lot d'une valeur de $a\$$ avec une probabilité x (et de ne rien gagner avec une probabilité $1 - x$) a comme espérance mathématique le produit $a\$ \times x$. En clair, l'espérance mathématique mesure adéquatement la valeur du pari puisque l'intérêt du lot diminue avec la probabilité de le gagner. Le degré de croyance,

est introduit comme un paramètre jouant le même rôle dans le même contexte que la probabilité pour le joueur :

> [...] ce qui revient à dire que si *p* est une proposition qu'il juge douteuse, tous les biens ou les maux pour lesquels la réalisation de *p* est une condition nécessaire et suffisante vont entrer dans ses calculs en tant que multipliés par la même fraction qu'on appelle « le degré de croyance de *p* ».[40]

Pour illustrer le concept de croyance partielle qu'il propose, Ramsey propose un exemple particulier où l'agent doit décider entre deux chemins à prendre et où, ayant choisi un des chemins et s'étant engagé sur une route, il a la possibilité de se dérouter pour aller demander à un informateur s'il a fait le bon choix. La distance qu'il devrait parcourir pour vérifier s'il a choisi la bonne route sert ici de mesure du degré de confiance qu'il a d'avoir choisi la bonne route. Il est clair qu'on ne peut généraliser la structure de cet exemple car il n'est pas toujours possible pour un agent d'obtenir une telle confirmation, sauf peut-être à attendre l'avènement des conséquences espérées. Appelons *condition particulière* cette supposition à l'effet que l'on peut estimer le coût associé au fait de confirmer ou d'infirmer la vérité de la proposition *p*. En gardant à l'esprit cette condition particulière, on peut examiner la formule pour le calcul de la croyance partielle. En respectant la notation utilisée par Ramsey, on interprète la formule suivante, où tous les paramètres sont relatifs à un agent fixe quelconque,

$$Cr^{\circ}(p) = 1 - \frac{f(d)}{r - w}$$

comme indiquant la valeur du degré de croyance partiel en la proposition *p*, où *f(d)* —pour une distance *d* — représente le coût pour cet agent de découvrir la valeur de vérité de la proposition *p*, *r* représente l'avantage pour lui d'avoir fait le bon choix et *w*, l'inconvénient d'avoir fait le mauvais choix. Les propriétés de la fonction ainsi définie permettent de prédire, conformément à l'intuition, que plus le degré d'assurance de l'agent dans son choix sera grand, moins il sera disposé à payer un coût élevé pour confirmer ou infirmer la vérité de *p*. En laissant de côté la condition particulière indiquée plus haut, on retrouverait le cas général où il n'est pas possible de déterminer la valeur de vérité de *p* autrement qu'en fin de compte, comme dans une loterie, lors du tirage[41]. On retrouverait la formule du ratio équitable d'un pari (*fair betting quotient*) dérivée de la formule de l'espérance mathématique et des axiomes de la probabilité.

Ramsey est pleinement conscient que cette définition de la croyance partielle camoufle un présupposé caché, celui de la neutralité à l'égard du risque[42]. Nous savons aujourd'hui que la disposition à accepter un risque varie selon les personnes. Ce fait, établi empiriquement par les recherches en psychologie, nous force à considérer trois possibilités : La valeur acceptable pour *f(d)* sera plus élevée si l'individu est disposé au risque (*risk prone*), moins si l'individu a une aversion pour le risque (*risk aversive*) et cette valeur ne sera égale au ratio équitable que pour un sujet qui est neutre à l'égard du risque. L'individu disposé au risque acceptera un pari qui lui est défavorable, celui qui a une aversion du risque refusera un pari qui l'avantage et seul celui qui est neutre à l'égard du risque révélerait par son choix le degré de croyance qu'il place dans la proposition *p*. Une solution à cette difficulté serait d'en connaître davantage sur la fonction d'utilité de l'agent avant d'observer son

comportement d'acceptation ou de refus d'un pari. Cependant, l'idée de procéder à l'interrogation du sujet est exclue dans la perspective de la méthodologie pragmatiste de Ramsey puisque cette interrogation ferait intervenir une forme d'introspection psychologique dans l'évaluation du degré de croyance. La solution de Ramsey est une innovation et une caractéristique importante de sa théorie. Elle consiste à calibrer l'échelle de préférence à partir de propositions dites « éthiquement neutres ».

On vient de voir qu'on peut utiliser le concept de degré de croyance et le concept de pari pour construire une interprétation des probabilités. Avant de passer à la construction de l'échelle de mesure des degrés de croyance, on doit faire quelques remarques pour préciser les concepts de croyance et de proposition. On a utilisé les termes « croyance » et « proposition » comme si le sens de ces termes était clair dans le langage philosophique alors qu'on sait que ces concepts sont problématiques. En formulant la théorie de Ramsey, on doit comprendre qu'une croyance n'est pas simplement une attitude propositionnelle — une disposition à tenir pour vraie la proposition qui représente un certain état de chose. Ramsey a proposé dans un autre texte une théorie relationnelle de la croyance qui est semblable mais différente de la théorie de Russell[43]. Une telle théorie associe à l'énoncé « Othello croit que Desdémone aime Cassio » la forme logique d'une relation à plusieurs termes

Cr (Othello, Desdémone, aime, Cassio).

Dans *The Nature of Propositions*, Ramsey [1921], Ramsey affirme que les propositions n'existent pas et que dans p & q, p et q réfèrent respectivement à « est une croyance que p » et « est une croyance que q ». Cette théorie de Ramsey [1921] est intéressante et originale mais ce n'est pas celle qui sera utilisée dans *Truth and Probability*, Ramsey [1926]. Ici, une croyance est

surtout une disposition à se comporter comme si la proposition qui constitue l'objet de la croyance était vraie en essayant de préserver une certaine cohérence entre les indicateurs verbaux et non-verbaux de cette croyance [44]. Pour Ramsey, la croyance est un concept fondamentalement pragmatique, tourné vers l'action. Voyons maintenant le concept de proposition. Nous devons tenir compte du fait que la théorie des propositions de Ramsey a évolué rapidement et qu'il s'est progressivement détaché de la philosophie du *Tractatus* de Wittgenstein. Il y a une affirmation centrale dont nous devons tenir compte et c'est l'indication donnée par Ramsey dans une note de *Truth and Probability.* Il y affirme « qu'il adopte la théorie de Wittgenstein en faisant l'hypothèse que la définition qu'il veut construire pourrait s'exprimer dans toute autre théorie des propositions »[45]. Nous avons remarqué que la plupart des auteurs en logique de la décision utilisent le concept de proposition sans le préciser ou en faire une discussion critique. Ils présupposent sans doute que ce concept est susceptible d'une élaboration indépendante ou ils ont en tête un concept opératoire et minimal qui serait implicite, par exemple, dans n'importe quel exposé du calcul des propositions. Comme nous venons de le voir, Ramsey est un peu plus explicite. Wittgenstein a indiqué lui-même dans l'introduction des *Investigations Philosophiques* que ce sont principalement les critiques de Ramsey qui l'ont conduit à reconnaître les erreurs du *Tractatus Logico-Philosophicus*. On peut affirmer sans la moindre hésitation que Ramsey était conscient des difficultés inhérentes au concept de proposition et on ne peut lui reprocher d'être évasif sur ces questions. Examinons quelques caractéristiques du concept de proposition chez Wittgenstein et Ramsey en tenant compte des autres écrits de Ramsey, en particulier Ramsey [1921], Ramsey [1927], mais aussi Ramsey [1929]b.

Wittgenstein a proposé dans le *Tractatus Logico-Philosophicus* une théorie des propositions qui veut répondre à la question « comment est-ce possible pour une combinaison de mots de représenter un fait dans le monde ? ». Pour répondre à cette question, il propose la théorie que la proposition est une image (*ein Bild*) de la réalité[46]. Cette théorie a comme conséquence que les constituants des propositions atomiques sont des noms car les constituants d'une image doivent aussi correspondre à des constituants du fait qu'elle indique. Ne nous attardons pas sur la composition interne des propositions car elle n'est pas utilisée ici. Pour reprendre l'analogie de Ramsey, pour étudier le jeu d'échecs, nous n'avons pas à décrire les atomes qui composent les pièces du jeu[47]. Les idées principales du concept opératoire de proposition chez Ramsey sont données dans Ramsey [1927]. Les unités de base sont les *propositions atomiques*. Elles peuvent être les constituants de propositions complexes mais elles n'ont pas elles-mêmes de constituants propositionnels. Les propositions peuvent être vraies ou fausses et sont, en logique, les porteurs des valeurs de vérité. De plus, les propositions atomiques peuvent être vraies ou fausses séparément les unes des autres. Ainsi, nous dit Ramsey, si j'ai n propositions atomiques, j'aurai 2^n possibilités mutuellement exclusives[48]. Ces sont les *possibilités de vérité* (*truth-possibilities*) qui peuvent servir à représenter les attitudes (ce que l'agent croit, ne croit pas[49], suppose,...). Nous utiliserons l'expression monde possible pour désigner ce que Ramsey appelle « possibilités de vérité ». Quoiqu'il soit possible de le faire, nous croyons qu'il est préférable de ne pas définir le concept de monde possible mais de le prendre comme terme primitif. Ramsey poursuit en soulignant que dans une langue naturelle, on peut toujours décrire une telle attitude par « le biais de la croyance en un énoncé compliqué formé de

propositions atomiques à l'aide de plusieurs conjonctions »[50]. Nous définirons plus loin le concept de proposition caractéristique qui précise cette idée. Les propositions caractéristiques sont des expressions, des formes syntaxiques qui représentent les mondes possibles, plutôt que des définitions. À l'aide des constantes logiques, il est possible de former des propositions moléculaires (négations, disjonctions). Ces propositions sont des fonctions de vérité des propositions élémentaires. La vérité ou la fausseté des propositions moléculaires est déterminée de façon univoque lorsque la valeur de vérité des propositions élémentaires qu'elle contient est fixée. Certaines de ces propositions moléculaires, parce qu'elles n'éliminent aucune possibilité, ne vont correspondre à aucune croyance. C'est le cas de « *p* ou non-*p* », et de toutes les autres *tautologies*. « Elles laissent ouvert tout l'espace logique » dit Ramsey. De même, certaines propositions moléculaires « excluent toutes les possibilités » comme « *p* et non-*p* » et n'expriment aucune attitude possible ; ce sont les *contradictions*. Pour Ramsey et Wittgenstein, ce sont des cas dégénérés qui ne sont l'image de rien. Vers 1927, Ramsey peut donc affirmer que sa théorie des propositions est identique à celle de Wittgenstein et ceci vaut également pour la structure des propositions générales, les propositions quantifiées. Cette adhésion complète à la théorie de Wittgenstein doit être nuancée sur un seul aspect. Ramsey suggère que le concept de proposition atomique est relatif à un langage, car certaines pensées complexes doivent être composées d'une combinaison d'états psychologiques (*feelings*) et de mots de telle manière que la spécification complète de ces pensées n'est possible que par référence à un langage déterminé[51]. Ceci complète notre discussion des concepts de croyance et de proposition dans le contexte de la théorie de Ramsey. Nous n'avons pas tenté de faire un examen complet des concepts de

proposition et de croyance, ni d'exposer ce qui nous semble être la meilleure explication de ces concepts, ni de répondre aux questions importantes qui se posent à leur sujet lorsqu'on cherche à construire un modèle de la délibération, de l'action et du choix rationnel. Plusieurs aspects de ces enjeux importants seront abordés dans les chapitres qui suivent mais le traitement détaillé et unifié de ces questions est une tâche ambitieuse qui fera l'objet de recherches ultérieures[52].

2.5 Les propositions éthiquement neutres

Il y a beaucoup à dire sur le concept de proposition éthiquement neutre car il constitue la clef de voûte de la construction de Ramsey. Ce concept permet de définir les valeurs numériques pour le concept de probabilité subjective à partir des préférences ou de l'indifférence du sujet pour certains mondes possibles qui sont les conséquences de choix qui se présentent comme des paris. Nous allons décrire les étapes de cette construction après avoir formulé les éléments de la problématique qui permettent d'aborder le concept de proposition éthiquement neutre de façon critique. Nous allons examiner successivement les questions de savoir si la théorie de Ramsey suppose un agent indûment idéalisé ou si l'atomisme logique inhérent à son système de définitions est recevable. Nous allons également discuter un raisonnement percutant dû à H. J. Sobel qui tente de démontrer qu'aucune proposition éthiquement neutre ne peut exister pour aucun sujet. Dans ces discussions critiques, il est important de garder à l'esprit que la théorie de Ramsey se propose de mettre en lumière les fondements des concepts de préférence et de degré de croyance. Dans d'autres contextes, il est possible d'accepter ces

concepts sans trop se soucier de la façon de les fonder. Ainsi, en d'autres occasions, on pourra simplement postuler que l'agent possède une fonction qui établit ses préférences selon un type d'ordre qui a les propriétés formelles souhaitables. De plus, l'approche fondationnelle de Ramsey n'est pas la seule. von Neumann et Morgenstern ont découvert indépendamment une théorie comparable à celle de Ramsey et leur traitement du concept « d'utilité » en est en quelque sorte une version simplifiée. Il est très clair que ces auteurs, comme De Finetti, Davidson et Suppes, ainsi que Savage, Ramsey ou Sobel interprètent les concepts élémentaires d'une logique de la décision fondée sur le concept de probabilité subjective de façons fort différentes. Dans la perspective logico-philosophique adoptée ici, on ne peut éviter les questions qui concernent l'acceptabilité des concepts et les choix théoriques concernant l'ontologie formelle. Selon nous, ces questions sont et doivent être de première importance.

Ramsey commence d'abord par reformuler le contexte théorique. Comme précédemment, on suppose un agent qui a une fonction de croyance bien définie et totale mais on met de côté l'hypothèse que les biens désirables sont additifs et immédiatement mesurables. L'agent va agir de façon à faire en sorte que les conséquences totales de ses actions soient les meilleures possibles, c'est-à-dire, selon la formule consacrée, de façon à en maximiser l'utilité espérée. Ainsi, les mondes possibles se trouvent ordonnés selon leur valeur, mais il n'est pas encore précisé comment leur associer des valeurs numériques. La relation de préférence-ou-indifférence est clairement réflexive, transitive et totale (connexe) sur l'ensemble des mondes et des options des paris[53]. Convenons de représenter par des lettres grecques minuscules les mondes possibles, que Ramsey identifie à des « totalités

organiques d'événements possibles »[54]. Si p est une proposition certaine, l'agent peut répondre directement aux questions de la forme « préférez-vous β si p est vraie ou γ si p est fausse ? ». En effet, il suffit de comparer la valeur des mondes en question pour faire un choix. Mais il s'agit à présent de formuler les principes et les axiomes gouvernant les choix dans le cas où p serait une proposition probable plutôt que certaine. Il est clair que l'agent peut associer une valeur à la réalisation de la proposition p elle-même, ce qui introduit une complication. C'est pour contourner cette difficulté et fonder le système qui permet de mesurer ces valeurs que Ramsey va introduire le concept de proposition éthiquement neutre.

définition 2.5.1 :

> On définit comme *éthiquement neutre* une proposition atomique p qui est telle que deux mondes possibles qui diffèrent au plus quant à la vérité de p sont toujours de valeur égale. D'une proposition qui n'est pas atomique, on dira qu'elle est éthiquement neutre si tous ses constituants atomiques vérifonctionnels sont éthiquement neutres.

Les mérites de ce concept peuvent être discutés sur le plan intuitif avant de clarifier son rôle dans la construction formelle. Ramsey lui-même nous invite à y voir une étape nécessaire pour contourner la difficulté évoquée plus haut. Puisque nous privilégions une interprétation normative de la logique de la décision, la légitimité de ce concept ne nous semble pas problématique. Il suffit de constater que l'inexistence de telles propositions n'est pas nécessaire et la cohérence interne du concept suffit à le légitimer. Mais, ceux qui revendiquent une interprétation descriptive de la logique de la décision s'interrogent sur l'existence réelle de telles propositions et la possibilité de les

construire. Ainsi, Sahlin accepte manifestement la question de savoir si de telles propositions existent comme une question pertinente. Dans le cadre de la psychologie expérimentale, il mentionne une série d'expériences rapportées dans une étude de Davidson, Suppes et Siegel et qui utilisent un dé dont les faces portent des inscriptions comme « ZOJ » et « ZEJ », soit des suites de lettres dont il a été établi expérimentalement qu'elles n'ont pas de significations ou de valeurs connotatives associées. Avec un tel dé, ces auteurs parviennent à construire un dispositif expérimental qui reproduit les conditions d'un choix à partir d'une proposition éthiquement neutre[55]. Pour ces auteurs, la question de l'existence réelle de propositions éthiquement neutres est une difficulté réelle, un test qui confirme la validité de ce concept. Pour sa part, comme nous avons déjà eu l'occasion de le remarquer, Ramsey n'entretient aucune illusion sur la valeur descriptive de son axiomatisation : il s'agit, écrit-il, d'une version « très schématique » de la situation dans la vie réelle[56]. Il ne fait aucun doute que Ramsey privilégie une interprétation normative de sa théorie. Cette différence nous rappelle l'importance de demeurer vigilant quant à ces enjeux méthodologiques.

On remarque également que la définition présuppose l'existence de propositions atomiques à la manière de la théorie des propositions de Wittgenstein. L'existence de telles propositions, c'est-à-dire de propositions dont la valeur de vérité est totalement indépendante de la valeur de toute autre proposition, peut être critiquée et elle est explicitement rejetée par Jeffrey[57]. Il s'agit d'une seconde difficulté d'ordre philosophique et elle est du même ordre que celle que nous venons d'examiner. Comme nous l'avons déjà souligné, Ramsey note au passage qu'on pourrait construire une définition semblable à la définition 2.5.1 en s'appuyant sur toute autre théorie des

propositions mais ce souhait semble difficile à réaliser. En particulier, il semble impossible d'éliminer entièrement l'atomisme logique. C'est ce qui ressort d'un examen soigneux des variations possibles de la théorie de Ramsey. Ainsi, H. J. Sobel soutient que la définition de Ramsey présuppose nécessairement l'élément problématique de l'atomisme logique, mais qu'il s'agit d'un atomisme logique « mince » ou minimal[58]. De plus, comme le note Sahlin, Ramsey en viendra à renoncer en partie à la théorie des propositions de Wittgenstein quelques années plus tard[59]. Il nous incombe de nous prononcer sur l'acceptabilité du concept de proposition éthiquement neutre et sur l'acceptabilité du concept de proposition atomique qui apparaît dans le *definiens*. Notre position découle directement de notre choix philosophique et méthodologique de séparer autant que possible les questions logiques des questions épistémiques et de séparer celles-ci des questions plus générales concernant le réalisme psychologique des concepts utilisés dans la reconstruction formelle du processus de choix rationnel. Notre intention n'est pas de nier l'importance des questions de ce type. Nous croyons cependant qu'il est surtout approprié d'admettre et de traiter de ces questions externes relatives au réalisme psychologique en relation au système construit considéré comme un tout. En particulier, il est plus approprié de les aborder dans la discussion des théorèmes de représentation et dans la discussion des conséquences de l'argument connu sous le nom de « dutch book »[60]. Dans une certaine mesure, sur le plan logique, on peut se réclamer de ce que Carnap appelle le principe de *tolérance* qui nous autorise à construire un système formel en se donnant les concepts et les instruments formels nécessaires à notre visée théorique[61]. On comprend fort bien que le principe de Carnap est d'abord énoncé à propos de questions qui relèvent de la syntaxe et nous lui

donnons une portée méthodologique plus générale, suivant en cela l'orientation philosophique générale de Carnap[62]. La question difficile est celle de savoir si on peut accorder à ces concepts le statut de *termes théoriques*. Si c'est le cas, on peut les considérer comme des fictions commodes, des variables qui font l'objet d'une quantification existentielle, pour reprendre l'explication des termes théoriques que Ramsey propose dans un autre essai[63]. Il n'y a pas de doute que l'atomisme logique constitue une méthode efficace pour distinguer les mondes possibles et indique, de façon générale, comment les engendrer. Il est acceptable comme une approximation et il est utilisé parce qu'on se concentre sur l'adéquation formelle qui concerne principalement les propriétés de cohérence et d'exactitude. Pour les raisons déjà évoquées, on doit reporter le traitement des questions qui concernent l'adéquation matérielle au sens large, incluant le problème du réalisme psychologique et les questions philosophiques d'ordre ontologiques[64].

Nous sommes maintenant en mesure d'examiner un raisonnement dû à Sobel qui s'attaque directement à la cohérence de la définition de proposition éthiquement neutre. Nous reconstruisons ce raisonnement à partir de l'explication de Sobel et nous adaptons l'exemple qu'il utilise[65]. La première prémisse du raisonnement est la définition elle-même, c'est-à-dire, la première clause de la définition 2.5.1 : Une proposition est éthiquement neutre si deux mondes possibles qui diffèrent au plus quant à la vérité de cette proposition sont toujours d'égale valeur. La seconde prémisse demande d'accepter que pour toute proposition, il existe un monde possible dans lequel je vais parier une certaine somme sur la vérité de cette proposition. Cette hypothèse est recevable puisque la définition de Ramsey porte sur tous les mondes possibles sans restriction. La troisième prémisse demande de considérer la proposition

suivante, qui est proposée comme un bon exemple de proposition éthiquement neutre au sens intuitif du terme, c'est-à-dire, une proposition dont la valeur de vérité n'importe à personne : « Le nombre de cheveux sur la tête du Premier ministre du Canada est un nombre pair ». Sur cette base, on considère en vertu de la seconde prémisse qu'il existe bien un monde possible dans lequel la vérité de cette proposition m'importe et ce monde est celui où j'ai parié une certaine somme sur la vérité de cette proposition. Un tel monde existe en vertu de la seconde prémisse. Par conséquent, ni cette proposition, ni aucune autre que je pourrais considérer n'est éthiquement neutre pour moi au sens de Ramsey, c'est-à-dire au sens de la première prémisse. Comme le raisonnement n'utilise aucune propriété qui me soit propre, on peut généraliser cette conclusion. Il s'ensuit qu'aucune proposition éthiquement neutre ne peut exister pour personne.

À première vue, cette réfutation de la première clause de la définition 2.5.1 semble très efficace dans sa brillante simplicité. Elle ne devrait cependant pas nous convaincre. En effet, le raisonnement nous demande de considérer que la proposition à l'égard de laquelle je semblais indifférent ne me sera pas indifférente dans certains mondes possibles où j'ai parié sur sa valeur de vérité. Mais alors, ces mondes ne sont pas distincts *au plus* par le fait que dans un de ces mondes, le nombre de cheveux sur la tête du Premier ministre est pair alors que dans l'autre, le nombre de cheveux sur la tête du Premier ministre est impair. Ils sont distincts aussi par le fait que dans l'un de ces mondes, j'obtiendrai le gain associé au pari. Par conséquent, la construction du contre-exemple de Sobel contredit l'atomisme logique contenu dans la définition. Les deux mondes considérés ne sont plus distingués uniquement par la valeur de vérité de la proposition en question

contrairement à ce qu'exige la définition. Ainsi, au lieu de constituer une réfutation de la définition d'une proposition éthiquement neutre, le raisonnement de Sobel s'attaque plutôt à la plausibilité de l'atomisme logique qu'elle contient ou à l'absence de restriction sur l'ensemble des mondes possibles auxquels il est fait référence. Une solution efficace considérée par Sobel serait de reformuler la définition en la restreignant à un quadruplet de mondes possibles. Cet amendement peut être considéré comme une révision nécessaire pour ceux qui acceptent l'objection de Sobel.

À la lumière de la discussion qui précède, on peut se prononcer sur l'acceptabilité de la définition de la proposition éthiquement neutre. Nous croyons avoir montré que sa légitimité est suffisamment robuste. Mais, tout bien considéré, notre prise de position doit être provisoire et nuancée en fonction des difficultés liées au concept de proposition atomique. Il appert que dans le contexte de la logique de la décision, conçue comme une reconstruction formelle de la théorie du choix rationnel, il s'avère difficile ou inapproprié de séparer entièrement les questions logiques et les questions doxastiques. Cette position semble être celle de Jeffrey qui rejette l'idée de proposition atomique. Tant que l'on considère les propositions dans leur fonction logique, en tant que porteurs de valeurs de vérité, le concept de proposition atomique est intelligible. Mais si les propositions sont conçues comme la sorte de chose qui peut être connue, crue, confirmée ou réfutée, si elles sont l'objet véritable des choix et des paris, l'idée de proposition atomique est moins attrayante. L'élaboration d'une logique de la décision exige la résolution de cette difficulté et de plusieurs autres difficultés de même nature. Comme nous reviendrons plus loin sur cette question spécifique pour proposer un traitement plus détaillé, nous nous limitons ici à formuler notre

position de façon sommaire. Nous reconnaissons la difficulté de considérer les conditions de vérité d'une proposition comme entièrement indépendantes des conditions de vérité d'autres propositions qui en sont les antécédents logiques ou les conséquences. Cette difficulté peut être reconnue sans qu'il soit nécessaire de souscrire à une épistémologie holistique radicale qui ferait dépendre la vérité de chaque proposition de la vérité de tout l'ensemble du système du savoir. Heureusement, nous croyons qu'il est possible de contourner cette difficulté dans l'élaboration d'une logique de la décision. Pour Ramsey, les croyances ont comme type logique celui des propositions tandis que les préférences ont comme type celui des conséquences, c'est-à-dire les mondes possibles qui résultent des choix. En réalité, dans les axiomes de Ramsey, les lettres grecques minuscules peuvent être interprétées tout aussi bien comme des mondes possibles que comme des valeurs au sens de l'échelle des préférences. Au chapitre VI, nous allons proposer que les propositions atomiques soient intégrées à des constructions logiques plus complexes représentant des faits momentanés actuels qui se produisent par suite d'actions des agents[66]. Je vais soutenir que les constructions logiques qui représentent ces états du monde sont de meilleures représentations des objets des choix que les propositions atomiques. Les questions d'ontologie formelle sont complexes et il est difficile de les aborder sans en fournir une formulation détaillée. Pour cette raison, la discussion des diverses possibilités est reportée au chapitre VI qui esquisse le contexte formel que nous jugeons approprié. Il semble que notre approche soit compatible avec la théorie de Ramsey et elle est indépendante du fait que la théorie de Ramsey soit statique, c'est-à-dire, qu'elle suppose que les croyances et les valeurs ne changent pas avec le temps. La prochaine étape de la construction de Ramsey consiste à définir le

degré de croyance 1/2 pour la proposition éthiquement neutre. L'idée est intuitivement simple ; il s'agit de décrire la situation de l'agent pour qui un pari sur *p* a exactement la même valeur qu'un pari sur *non p*.

définition 2.5.2 :

> On dit d'un sujet qu'il a une croyance de degré 1/2 en une proposition éthiquement neutre s'il n'a pas de préférence entre les deux options représentées dans le tableau suivant, bien qu'il ait une préférence entre α et β:

	option 1	option 2
si *p* est vraie	α	β
si *p* est fausse	β	α

Voyons ce qu'indique ce tableau à l'aide d'un exemple. On dit d'une personne qu'elle juge qu'il est aussi probable qu'il neige ou qu'il ne neige pas (dans l'option 1) si c'est indifférent pour elle de gagner cinq dollars (la conséquence α) s'il neige et de perdre de cinq dollars (la conséquence β) s'il ne neige pas ou (dans l'option 2) d'obtenir une perte de cinq dollars s'il ne neige pas et un gain de cinq dollars s'il neige.

Deux conditions importantes viennent s'ajouter à la définition 2.5.2. En premier, on suppose que les gains représentés par α et β sont compatibles avec la vérité ou la fausseté de la proposition *p* comme c'est le cas dans notre exemple. En second lieu, un axiome « sans numéro » vient préciser que s'il y a indifférence entre les options 1 et 2 avec la paire de gains possibles α et β, il

en sera de même pour toute autre paire de gains possibles. Il s'agit d'un axiome de consistance ayant un contenu normatif pour un agent idéalisé.

La définition 2.5.2 sert de base à la construction de l'échelle qui mesure la préférence. Le concept de différence de valeur entre α et β est précisé lorsqu'on a indiqué le sens de l'expression « la différence de valeur entre α et β est la même que la différence de valeur entre γ et δ ». Or le sens de cette expression est défini à partir du tableau précédent en effectuant les remplacements appropriés dans la seconde ligne. Ceci nous donne la définition principale 2.5.3 *de l'égalité de la différence de valeurs entre deux paires de mondes* dont la clause est représentée dans le tableau qui suit.

définition 2.5.3

Soit un agent quelconque et une proposition *p* qui est éthiquement neutre et à laquelle cet agent accorde un degré de croyance 1/2. On dit que la différence de valeur entre α et β est la même que la différence de valeur entre γ et δ si et seulement si l'agent n'a pas de préférence entre les deux options représentées dans le tableau suivant, bien qu'il ait une préférence entre α et β:

	option 1	option 2
si *p* est vraie	α	β
si *p* est fausse	δ	γ

La dernière étape consiste à formuler de façon générale la valeur associée à un monde. Elle se définit comme la classe de tous les mondes qui lui sont également préférables.

> On suppose que si un monde α est préférable à β, tout monde ayant la même valeur que α sera préférable à tout monde qui a la même valeur que β et nous dirons que la valeur de α est plus grande que la valeur de β.[67]

L'approche utilisée par Ramsey dans ces définitions correspond à une stratégie habituelle en logique qui consiste à dériver les définitions des expressions analysées à partir de leur critère d'identité. Ainsi, le concept de valeur est défini à partir du sens de l'expression « avoir la même valeur » et le concept de différence de valeur est obtenu à partir de la définition du concept de « différence égale de valeur ».

2.6 Illustrations des définitions précédentes

Dans cette section, nous allons illustrer de quelques exemples les relations que permettent d'exprimer la théorie de Ramsey. On retrouvera des exemples comparables dans Sahlin [1990], Jeffrey [1965], Jeffrey [1965]b ainsi que Sobel [1998]. Il faut prendre note que la notation utilisée par ces auteurs est parfois différente de celle de Ramsey dans la mesure où ils représentent les même notions en utilisant les termes primitifs de leurs propres théories pour représenter celle de Ramsey. Par contraste, nous nous sommes efforcés de rester au plus près de l'appareil théorique développé dans Ramsey [1926]. Nos illustrations veulent mettre en évidence les relations

conceptuelles que la théorie de Ramsey permet de formuler et qui ne sont pas formulés explicitement dans son article.

Nous avons déjà proposé un exemple pour le concept de proposition éthiquement neutre dans la discussion de l'objection de Sobel. Reprenons-le en supposant cette fois qu'il s'agit bien d'une proposition éthiquement neutre au sens de Ramsey.

p = Le nombre de cheveux sur la tête du Premier ministre du Canada est un nombre pair.

Un pari sur cette proposition peut s'écrire de la façon suivante (α ;p ;β) où, suivant la notation de Ramsey, αreprésente le monde possible qui constitue la conséquence si p est vraie et β représente le monde possible qui constitue la conséquence si p est fausse. Pour les fins de l'illustration, interprétons respectivement α et β comme un gain de 5$ et une perte de 5$. Ainsi, la formule (5$; p ; –5$) représente le pari où nous gagnons cinq dollars si le nombre de cheveux sur la tête du Premier ministre est un nombre pair et où je perds cinq dollars dans le cas contraire.

Introduisons la relation de préférence, notée $\succ$, pour abréger « ...est préféré à... » ou « ...est mieux que... ». Pour la plupart d'entre nous, (5$ $\succ$ – 5$). Si nous n'avons aucune préférence pour le pari (5$; p ; – 5$) et le pari (– 5$; p ; 5$), c'est le signe que nous accordons une probabilité égale p et à $\neg p$. Nous venons ainsi de dériver le degré de croyance en p à partir de l'indifférence entre deux paris.

En utilisant la relation de préférence, on peut formuler une clause qui correspond à la condition nécessaire et suffisante pour la proposition éthiquement neutre. Cette clause s'écrit de la façon suivante[68]:

$\neg(\alpha \,\&\, p \succ \alpha) \,\&\, \neg(\alpha \succ p \,\&\, \alpha)$

En interprétant α et p comme précédemment, la paraphrase de cette formule énonce que la proposition que le nombre de cheveux sur la tête du Premier ministre du Canada est un nombre pair est une proposition éthiquement neutre parce que je n'ai pas de préférence entre d'une part, gagner 5 dollars et que la proposition soit vraie et d'autre part gagner 5 dollars.

Introduisons maintenant la notation $q1/2$ pour désigner une proposition q quelconque pour laquelle un agent donné a le degré de croyance de 1/2. De même, on introduit la relation d'indifférence « ...n'est ni mieux ni pire que... », notée $\approx$. La forme $(\gamma \approx \delta)$ peut se définir de la façon suivante :

$(\gamma \approx \delta) =_{\text{déf}} \neg(\gamma \succ \delta) \,\&\, \neg(\delta \succ \gamma)$

La notation introduite pour l'équivalence de paris peut être utilisée pour représenter les clauses de la définition d'une proposition éthiquement neutre dont le degré de croyance est 1/2 (la définition 2.5.2). Ces clauses peuvent s'écrire de façon abrégée de la façon suivante: $(\alpha; q1/2\ ;\ \beta) \approx (\beta; q1/2\ ;\ \alpha)$. Cette formule exprime que les deux paris sont jugés également avantageux.

En interprétant α et β comme précédemment, $(\alpha = +5\$,\ \beta = -5\$)$ [69], la proposition est vraie et rationnelle pour un agent quelconque. On peut utiliser cette formule pour exprimer sous forme abrégée l'égalité de la différence entre deux paires de mondes introduite par la définition 2.5.3. Cette définition permet de préciser l'idée de commensurabilité des préférences.

définition 2.5.3 (reformulée)

(Selon les conditions et conventions d'écriture que nous venons d'expliquer) la différence de valeur entre α et β est la même que la différence de valeur entre γ et δ si et seulement si
$(\alpha; q1/2; \delta) \approx (\beta; q1/2; \gamma)$.

Notre dernière tâche dans cette section est de montrer comment relier la définition 2.5.3 à la formule générale qui explique le choix rationnel par la maximisation de l'utilité espérée. Pour cette illustration, nous allons supposer que la fonction d'utilité $U(w_i)$ de l'agent est connue. Pour Ramsey, chaque action peut être vue comme le choix d'une option dans un pari[70]. Nous voulons maintenant préciser la valeur subjective, la désirabilité ou ce qu'on appelle *l'utilité espérée* d'un pari de la forme $(\alpha ; p ; \beta)$. Nous allons examiner un choix très simple mais il faudra surmonter quelques complications formelles pour l'aborder d'une façon précise, en toute généralité et en respectant au plus près l'approche de Ramsey.

Considérons la valeur $V(A_n)$ pour un acte A_n. Les actes qui forment l'ensemble A_1, A_2, A_3,..., A_n sont les actes que l'agent peut faire dans les circonstances du choix ; ce sont ceux à propos desquels il délibère. On se souviendra que pour Ramsey, une action est représentée par le monde (l'état des choses) qui résulte du fait d'accomplir cette action. Bien que ce ne soit pas dit explicitement, on comprend qu'il s'agit d'une action *type* et non d'une action *token*. On doit cependant ajouter que cette action-type est finement distinguée des autres. Si je considère qu'il est préférable d'aller nager plutôt que de continuer à lire à ce moment-ci ou je décide d'interrompre ma lecture et d'aller nager, ce n'est pas l'action de nager en général qui est évaluée par la fonction $V(A_n)$ mais l'action, aussi finement individuée que nécessaire, qui

consiste pour *moi* à aller nager *maintenant*, avec tous les éléments distinctifs qui sont pertinents pour définir la circonstance présente telle que distinguée de toute autre circonstance où je pourrais évaluer différemment l'*utilité* d'aller nager. De plus, on doit tenir compte de la contrainte de type spécifique à la théorie de Ramsey pour l'argument de la fonction de probabilité. Pour inscrire ceci dans la syntaxe, nous devons construire et donner un nom à un énoncé qui représente chaque monde possible. Plus précisément, on introduit un terme propositionnel qui dénote chacun des éléments de l'ensemble des mondes possibles qui sont accessibles dans les circonstances de la délibération de l'agent[71]. Comme David Lewis, nous trouvons préférable ici de ne pas définir le concept d'un *monde possible* mais plutôt de le prendre comme terme primitif[72]. Pour cette raison la proposition caractéristique d'un monde que je construis n'est pas une *définition stricto sensu* de ce monde possible mais simplement une représentation linguistique de ce dernier[73]. Dans une théorie comme celle de Ramsey, les mondes sont uniquement déterminés par les propositions atomiques qui sont vraies à ces mondes. La proposition caractéristique d'un monde selon un modèle est la conjonction formée de toutes les propositions atomiques qui sont vraies à ce monde selon ce modèle ainsi que des négations de chaque proposition atomique qui est fausse à ce monde selon ce modèle. La proposition caractéristique ainsi construite est maximale au sens où l'ajout de tout autre terme propositionnel — atome ou négation d'atome— formerait une proposition inconsistante. Cette maximalité peut être utilisée pour définir la proposition caractéristique d'une monde. On dira que $w_i^{[+p]}$ dénote la proposition caractéristique du monde $w_i \in$ W ou p est vraie et qui est la conjonction

$$\ldots p \,\&\, q \,\&\, r \,\&\, \neg u \,\&\, \neg v \,\&\, \neg w \ldots$$

construite à partir de chaque proposition atomique ...p, q, r, vraies à w_i et de la négation de chaque proposition u, v, w, ...qui sont fausses à w_i de telle façon que pour tout terme propositionnel s,
ou bien s est le n^{ième} terme de $w_i^{[+p]}$
ou bien $\neg s$ est le n^{ième} terme de $w_i^{[+p]}$
ou alors le résultat d'ajouter s (ou d'ajouter $\neg s$) à $w_i^{[+p]}$ rendrait $w_i^{[+p]}$ inconsistante[74]. Comme c'est l'usage, les points de suspension indiquent la possibilité d'éléments additionnels, ici dans les deux directions. Ainsi, chaque proposition $w_i^{[+p]}$ est vraie d'un seul monde. Sur cette base, on introduit le terme Prob ($w_i^{[+p]}$) qui dénote une fonction de probabilité dont l'argument sera la proposition caractéristique du monde où le résultat w_i sera obtenu si p est vraie pour chaque w_i qui pourrait arriver si p est vraie. On définit Prob ($w_i^{[-p]}$) de la même façon, l'argument de la fonction étant $w_i^{[-p]}$ si p est fausse. Pour simplifier et en idéalisant la situation, Ramsey nous demande de supposer que l'agent n'a pas de préférence à l'égard de la vérité ou de la fausseté de p et que le pari est proposé par un être tout-puissant qui est en mesure de réaliser w_i si p est vraie.

$$V(A_n) = \sum_i U(w_i) \times \text{Prob}\,(w_i)^{[p]}$$

Dans cette formule générale, le terme $w_i^{[p]}$ est utilisé (sans signe devant le p) parce que la valeur de p n'est pas connue au moment du pari. L'idée est intuitivement claire, mais pour la rendre entièrement explicite, on doit préciser les conditions de vérité de $w_i^{[p]}$. Comme indiqué, l'argument de la fonction de probabilité doit être une proposition plutôt qu'un monde. De plus, il faut tenir compte du fait qu'au moment d'évaluer un pari, la valeur de vérité de la

proposition *p* dans le pari (α ; p ; β) n'est pas encore déterminée. C'est pourquoi nous allons suivre Sobel qui, en formulant la théorie de Ramsey, introduit le concept de quasi-mondes (*near-worlds*), notés $w^{[p]}$. Ils se représentent comme la disjonction $w^{+p} \vee w^{-p}$ formée à partir des conjonctions propositionnelles qui caractérisent les mondes où respectivement *p* est vraie et *p* est fausse[75]. Les termes de cette disjonction sont définis de la façon suivante:

Pour α^{+p},

si *p* est vraie dans α, ce terme dénote la proposition caractéristique du monde α;

si *p* est fausse dans α,ce terme dénote la proposition caractéristique du monde β qui est tel que pour toute proposition *q*, *q* est vraie dans β si et seulement si *q* est vraie dans α ou *q* est identique à *p*.

Pour α^{-p},

si *p* est fausse dans α, ce terme dénote la proposition caractéristique du monde α;

si *p* est vraie dans α,ce terme dénote la proposition caractéristique du monde β qui est tel que pour toute proposition *q*, *q* est vraie dans β si et seulement si *q* est vraie dans α et *q* est distincte de *p*.

Pour une paire composée d'un monde *w* et d'une proposition atomique *p*, on définit la proposition caractéristique du quasi-monde $w^{[p]}$ comme la proposition $w^{+p} \vee w^{-p}$. Le terme $w^{[p]}$ pour le *quasi-monde* fondé sur *p*, est une simple abréviation de cette disjonction.

On dispose maintenant de la formule générale qui permet de calculer la valeur d'un acte et nous allons l'utiliser pour comparer des paris. Pour mon illustration, les actes envisagés dans la délibération correspondent à l'acceptation d'un des paris suivants:

A_1 : (α; q1/2; β)

A_2 : (γ; q1/2; δ)

De plus, nous supposons que la fonction $U(w_i)$ donne les valeurs suivantes:

$U(\alpha) = +5\$$,

$U(\beta) = -5\$$,

$U(\gamma) = +7\$$,

$U(\delta)$ = *statu quo*, l'agent ne gagne rien et il ne perd rien.

Voyons comment le second pari sera préféré au premier, $A_2 \succ A_1$ pour un agent qui cherche à maximiser l'utilité espérée. La valeur du premier pari se calcule en évaluant

$$V(A_1) = U(\alpha) \times \text{Prob}(\alpha^{[+q1/2]}) + U(\beta) \times \text{Prob}(\beta^{[-q1/2]}).$$

$$= (+5\$ \times 1/2) + (-5\$ \times 1/2)$$

$$= 0\$$$

alors que la valeur du second pari sera

$$V(A_2) = U(\gamma) \times \text{Prob}(\gamma^{[+q1/2]}) + U(\delta) \times \text{Prob}(\delta^{[-q1/2]}).$$

$$= (+7\$ \times 1/2) + (0\$ \times 1/2)$$

$$= 3{,}50\$$$

Comme il se doit, $V(A_2) > V(A_1)$ et le second pari devrait être préféré au premier.

2.7. L'axiomatisation

Examinons les huit axiomes proposés par Ramsey. Certains de ces axiomes précisent des contraintes de rationalité, d'autres sont plutôt des axiomes structuraux qui garantissent certaines propriétés de l'échelle des valeurs. Il n'est pas facile de départager clairement les axiomes selon ces deux catégories. Typiquement, un axiome structural ne comporte aucune référence explicite ou implicite au concept de préférence et il concerne plutôt l'existence d'un ensemble ou de propositions d'un certain type. Nous indiquerons au passage ceux qui ont un contenu normatif et ceux qui ne sont pas indispensables pour une explication du choix rationnel. Les axiomes qui n'ont pas de contenu normatif doivent être considérés comme des postulats de l'approche fondationnelle de Ramsey. Comme nous l'avons déjà remarqué, il s'agit d'une axiomatisation partielle pour un concept fort de rationalité. On verra plus loin, à la section 2.9 du présent chapitre, dans quelle mesure ces axiomes sont censés être suffisants pour dériver les concepts nécessaires et comment ils sont nécessaires pour établir les résultats d'adéquation.

Axiome 1: Il y a une proposition p dont le degré de croyance est 1/2.

Comme nous l'avons expliqué dans la section 5 du présent chapitre, cet axiome est nécessaire pour l'approche fondationnelle de Ramsey mais on ne doit pas conclure qu'il est absolument nécessaire pour caractériser le choix rationnel dans un modèle de la délibération. Il est bien clair que cet axiome n'a pas de contenu normatif. De plus, comme Pfanzagl [1967] l'a montré, on peut

reformuler la théorie de Ramsey de façon à ne pas accorder un statut spécial à la proposition ayant un degré de croyance 1/2.

Les quatre axiomes qui suivent (2, 2a, 3, 4) sont des conditions nécessaires pour le choix rationnel. Ils imposent des normes à un agent idéal en posant certains paris comme équivalents dans une échelle de préférence qui respecterait la théorie de Ramsey. Rappelons que Ramsey utilise les lettres grecques minuscules pour désigner tantôt les mondes qui résultent des choix, tantôt pour désigner les valeurs dans ce que nous appelons l'échelle de préférence. Dans certains axiomes, elles doivent manifestement s'interpréter comme des valeurs, plus précisément, comme les valeurs de mondes ou de quasi-mondes. Il est important de remarquer que les axiomes concernent les distances entre ces valeurs, c'est-à-dire, des segments. Enfin, il faut rappeler cette particularité de la notation de Ramsey qui veut que l'expression $\alpha\beta$ dénote la différence de valeur (distance) entre α et β.

Nous allons suivre l'usage de Sobel et formuler l'axiome suivant à l'aide de la notation introduite pour les quasi-mondes.

Axiome 2: Si *p* et *q* sont des propositions éthiquement neutres, pour tout monde $\alpha, \beta, \gamma, \delta$, si

$(\alpha^{[p]}; p; \delta^{[p]})$ est équivalent à $(\beta^{[p]}; p; \gamma^{[p]})$,

alors

$(\alpha^{[q]}; q; \delta^{[q]})$ est équivalent à $(\beta^{[q]}; q; \gamma^{[q]})$

(2a) Si $\alpha\beta = \gamma\delta$, alors $\alpha > \beta$ si et seulement si $\gamma > \delta$

et $\alpha = \beta$ si et seulement si $\gamma = \delta$.

Définition: Dans le cas qui précède, on dira que $\alpha\beta = \gamma\delta$

On peut formuler l'axiome 2 en se concentrant sur la différence entre les deux paires de paris qui figurent respectivement dans l'antécédent et le conséquent de la conditionnelle. Cet axiome indique que si *p* est une proposition éthiquement neutre et que deux paris ayant pour objet la proposition *p* offrent des gains espérés jugés équivalents entre eux, alors les mêmes paris devraient être jugés équivalents s'ils ont pour objet la proposition *q*, pourvu que *q* soit également une proposition éthiquement neutre. Quant à l'axiome 2a, il énonce une propriété nécessaire pour l'échelle des préférences et il constitue une contrainte de cohérence pour l'agent.
Les deux axiomes qui suivent (3 et 4) concernent la transitivité.

Axiome 3: Si l'option A est équivalente à l'option B et que l'option B est équivalente à l'option C, alors l'option A est équivalente à l'option C.

Axiome 4: Si $\alpha\beta = \gamma\delta$, $\gamma\delta = \eta\zeta$, alors $\alpha\beta = \eta\zeta$.

Le sens de l'axiome 3 est évident. L'axiome 4 énonce pour la différence de valeur ce que l'axiome 3 énonce pour la préférence de certaines options. Gärdenfors et Sahlin font la remarque suivante en discutant spécifiquement l'axiome 4 mais comme ils le notent, elle s'applique à *tous* les axiomes,

> Remarquez que la formulation de cet axiome ne présume pas que nous ayons assigné quelque nombre que ce soit à la valeur des résultats, seulement que nous comparions les grandeurs (*magnitudes*) des distances entre les valeurs (de même pour les autres axiomes).[76]

Cette observation est nécessaire pour respecter l'ordre de la construction et ne pas utiliser ce qui n'a pas encore été introduit. Dans l'interprétation des axiomes qui vont suivre, pour fixer les intuitions par le chemin le plus facile, nous ne respecterons pas cette contrainte.

On ne peut guère discuter les axiomes 3 et 4 sans introduire quelques éléments pour une compréhension critique de leur légitimité. En effet, la transitivité de la relation de préférence est une propriété qui a été contestée par les critiques de la théorie de décision. Il est possible de construire des « contre-exemples » dans un test empirique où l'on demande au sujet s'il accepte les clauses qui forment l'antécédent lorsqu'elles sont considérées séparément (ou ensemble) pour constater ensuite qu'il refusera le conséquent. Il peut arriver aussi que nous ayons des intuitions fortes qui coïncident avec la réaction du sujet. Autrement dit, que nous trouvions cette réaction *raisonnable.*

On peut offrir comme exemple le cas de quelqu'un qui trouverait préférable de visiter Rome plutôt que de faire de la randonnée dans les Alpes mais qui trouve préférable de rester à la maison plutôt que de visiter Rome. Cependant, s'il avait le choix entre rester à la maison et faire de la randonnée dans les Alpes, il trouverait préférable de partir faire de la randonnée dans les Alpes.[77] Dans ce cas, il y a manifestement le facteur incident de considérer les choix deux à deux. En d'autres cas, il arrive que les clauses de l'antécédent n'apparaissent acceptables qu'en vertu d'un présupposé caché du genre « toutes choses étant égales par ailleurs » *(ceteris paribus)* ou en vertu d'un présupposé à l'effet que les options ou les résultats sont indépendants les uns des autres. Dans de tels cas, il peut sembler que la transitivité des préférences n'est pas

réalisée mais on peut aussi conclure de ces observations que, toutes choses bien considérées, la conjonction des clauses qui forment l'antécédent ne devrait pas être acceptée selon le concept de préférence privilégié par une théorie interprétée de façon normative[78]. En somme, on oppose à ceux qui remettent en cause la transitivité des préférences qu'un agent peut se tromper sur ses propres préférences, qu'il peut les comprendre imparfaitement, etc.[79]. Nous avons déjà soulevé ces difficultés pour signaler que ces axiomes, en particulier l'axiome 3, n'ont pas toujours été reçus comme des postulats évidents du choix rationnel. Nous traiterons à nouveau de ce thème aux chapitres IV et VI. Pour Ramsey, il est clair que ces axiomes sont des normes de rationalité plutôt que des axiomes structuraux. Nous n'avons pas le choix de les accepter si nous voulons être *rationnels*[80]. Mais on *pourrait* les considérer comme des axiomes structuraux qui énoncent des propriétés qui sont nécessaires pour la cohérence formelle de la structure de l'échelle de préférence dans la théorie de Ramsey.

Axiome 5: Pour tout α, β, γ, il y a un et un unique x tel que $(\alpha x = \beta\gamma)$

Intuitivement, cet axiome énonce l'existence et l'unicité d'une valeur x qui assure une propriété de commensurabilité pour les différences de valeur. Dans la formulation de cet axiome, nous avons pris la liberté d'exprimer par une paraphrase ce que Ramsey énonce à l'aide de la notation $\ldots\exists!(\iota x)\ldots$ des *Principia Mathematica* de A. N. Whitehead et B. Russell[81].

Les axiomes qui suivent (6, 7, 8) garantissent que l'ensemble des valeurs possède la structure des nombres réels. Ils ne sont pas des postulats nécessaires pour le choix rationnel: on pourrait construire une théorie du choix rationnel sans eux. Les formules que nous suggérons pour compléter

l'interprétation des axiomes suggérés par Ramsey sont construites sur le modèle des axiomes usuels pour les nombres réels et ce sont ceux qu'utilisent Fishburn [1990].

Axiome 6: Pour tout α, β, il y a une et seule valeur x telle que $(\alpha x = x\beta)$.

Intuitivement, cet axiome énonce que pour toute paire de monde, il en existe exactement un troisième dont la valeur est située à équidistance des valeurs des deux autres. Ainsi, on peut conclure à l'égalité des distances. Cette propriété (*compacité*) est aussi vérifiée par les nombres rationnels.

En formulant une axiomatisation pour les nombres réels, on spécifie habituellement la propriété suivante, clairement plus faible:

Soit **R**, l'ensemble des nombres réels, **Q**, l'ensemble des nombres rationnels.

Si $x, y \in \mathbf{R}$ et si $x < y$, il existe un $r \in \mathbf{Q}$ tel que $x < r < y$

Axiome 7: (axiome de continuité) Toute progression a une limite (ordinale).

Ramsey n'indique ni l'objet, ni la forme précise de l'axiome de continuité qu'il a en tête. Compte tenu de l'usage des concepts de "progression" et de "limite ordinale", on peut penser qu'il est vraisemblable que Ramsey se réfère à l'appareil conceptuel de Russell pour penser ces questions[82]. Un examen soigneux de ces possibilités est de grande importance pour ceux qui, comme Sobel, cherchent à formuler des extensions conservatrices de la théorie de Ramsey. Ceci n'étant pas notre but ici, nous proposons l'axiome suivant dont on sait qu'il est suffisamment fort dans le présent contexte (c.-à-d., avec les axiomes précédents). Il se fonde sur l'idée de coupure de Dedekind:

Soit A et B des ensembles non-vides, si $x \in \mathbf{R}$ et si $A = \{ a \mid a \in \mathbf{R} \ \& \ a < x \}$ et $B = \{ b \mid b \in \mathbf{R} \ \& \ b > x \}$, alors $\{ A, B, \{ x \} \}$ constitue une partition de $\mathbf{R}$.

Cet axiome n'est pas vérifié par les nombres rationnels. Il permet de déduire les propriétés caractéristiques usuelles concernant l'existence d'un supremum (infimum) pour les segments qui admettent un majorant (minorant).

Axiome 8: (axiome d'Archimède)

Cet axiome peut être formulé de différentes façons et ici encore, Ramsey n'indique pas la formule précise qu'il a en tête[83]. Considérons d'abord la forme habituelle que prend cet axiome en tant qu'il exprime une propriété des nombres réels, $\mathbf{Z}^+$ désignant les entiers positifs:

Si $a, b \in \mathbf{R}$ et si $a > 0$, il existe un $n \in \mathbf{Z}^+$ tel que $(n \times a) > b$

On sait que l'axiome d'Archimède pourrait aussi être formulé en termes de distances[84]. En regard de l'observation de Gärdenfors et Sahlin que nous avons citée en présentant l'axiome 4 et en tenant compte du rôle de l'axiome d'Archimède dans la construction, on peut affirmer que Ramsey *aurait* formulé l'axiome 8 en termes de distances. Ainsi formulé, il énonce que même si la distance entre α et β est infime et que la distance entre γ et δ est très grande, il existe un nombre entier tel que $(n \times \alpha\beta) \geq \gamma\delta$. De façon imagée, on peut dire qu'il sert à fabriquer l'étalon de mesure de l'échelle des valeurs. L'essentiel est que cet axiome a pour effet d'exclure les différences de valeurs de grandeur infinie.

Ceci complète notre examen des axiomes de Ramsey. Dans la prochaine section, nous verrons les principales conséquences théoriques que Ramsey peut établir sur cette base.

2.7 Définitions consécutives et propositions corollaires

Nous allons passer en revue les principales propositions qui découlent des axiomes de Ramsey. La première et peut-être la plus importante doit permettre d'établir une correspondance biunivoque entre les valeurs que nous venons d'introduire et les nombres réels. Il faut ajouter un qualificatif à cette identification à l'effet que cette correspondance (fonction) est unique seulement relativement à toutes les transformations linéaires positives (transformations affines positives)[85]. En conséquence, on a des classes d'équivalence et les valeurs ne sont pas distinguées une à une.

La proposition clef est la suivante où les termes $\alpha^1, \beta^1, \gamma^1, \delta^1$ sont les nombres réels qui correspondent respectivement aux valeurs $\alpha, \beta, \gamma, \delta$.

$$(\alpha\beta = \gamma\delta) \equiv ((\alpha^1 - \beta^1) = (\gamma^1 - \delta^1))$$

Ce résultat est central dans la théorie de Ramsey car il permet de définir le concept de croyance en général, c'est-à-dire, en anticipant un peu, de dériver le concept de probabilité subjective. Le degré de croyance en p, $Cr°$ (p), est introduit par la définition suivante où il est supposé que l'agent est indifférent entre l'option α si *p* est certain, et β si p est vraie et γ si *p* est fausse.

Dans cette alternative, on doit supposer que $p \in \beta$ et $\neg p \in \gamma$ et qu'il y a au moins un monde (d'une valeur quelconque) où *p* est vraie et un autre où *p* est fausse[86]. Il n'est plus requis que *p* soit éthiquement neutre.

définition 2.8.1:

$$Cr°(p) = \frac{(\alpha - \gamma)}{(\beta - \gamma)}$$

Le degré de croyance en la proposition *p* serait le même pour toutes les autres valeurs α, β, γou δ qui satisferaient les mêmes conditions. La définition 2.8.1 établit la relation entre le concept de degré de croyance et celui de pari.

Voici une situation qui illustre les composantes de la définition 2.8.1. Soit une loterie où l'on me propose de choisir un chiffre compris entre 1 et 6 inclusivement. Le coût pour participer à cette loterie est de 1 $. On lance un dé et si le chiffre choisi est le chiffre gagnant, je gagne une somme de 6 dollars. La probabilité de gagner est d'une chance sur 6. La relation entre le coût du billet, le lot à gagner et la probabilité de le gagner correspond à une cote équitable pour ce pari. En utilisant cet exemple pour faire un calcul à l'aide de la définition 2.8.1, nous obtenons le calcul suivant pour l'interprétation

p : la prédiction que le chiffre choisi sera gagnant

α: ce que j'obtiendrais certainement, qui sert ici de point de référence

γ: l'inconvénient de perdre, 0 $[87]

β: l'avantage de gagner, 6$

$$Cr°(p)=\frac{(\alpha-\gamma)}{(\beta-\gamma)}=\frac{\alpha-0}{6-0}=\frac{\alpha}{6}$$

Dans cet exemple, on voit que le degré de croyance en *p* — que le chiffre choisi est le bon — correspond à la probabilité de gagner le lot puisqu'en l'exprimant en termes de probabilité, α=1, la valeur du résultat certain.

Profitons de l'occasion pour introduire un peu plus de terminologie à propos des paris. Nous adaptons les définitions fort utiles proposées dans Earman [1992]. On a déjà introduit la forme (+a$, *p*, –b$). Un *pari* est une entente entre un parieur et preneur au livre (*bookie*). Selon cette entente, le parieur accepte de verser au preneur au livre la somme *b*$ (parfois notée –*b*$) si la proposition *p* qui fait l'objet du pari se révèle fausse et le preneur au livre

s'engage à verser au parieur la somme $a\$$ (parfois notée + $a\$$) si la proposition p est vraie. La somme $(a + b)\$$ s'appelle *les enjeux* (*stakes*) d'un pari sur p. La fraction b/a s'appelle la *cote* (*odds*) du parieur. En utilisant le degré de croyance plutôt que la probabilité, on peut définir la *valeur espérée d'un pari* sur p, pour le parieur, par la formule suivante:

définition 2.8.2:

Valeur espérée de (+a\$, p, –b\$) = $((a\$ \times Cr^{\circ}(p)) - (b\$ \times Cr^{\circ}(\neg p))$.

Un pari est réputé *équitable*, *favorable* ou *défavorable*, selon le cas, si sa valeur espérée est égale à zéro, positive ou négative. On arrive ainsi au concept de *ratio équitable d'un pari*, donné par la formule :

$$\frac{b\$}{(a\$ + b\$)}$$

Le ratio équitable d'un pari est ce que Ramsey appelle la « cote formulée en termes de différences » dans la citation suivante. En substituant les équivalents et en utilisant les lois de l'arithmétique, on démontre aisément que le ratio équitable d'un pari équivaut à la définition 2.8.1.

> Le degré de croyance de l'agent en p équivaut, en gros, à la cote (*odds*) pour laquelle il accepterait le pari s'il est formulé en termes de différence de valeurs selon la définition 2.8.1. [88]

Comme l'indique Ramsey, l'expression « en gros » dans la citation précédente annonce qu'il est nécessaire de formuler une réserve. La définition 2.8.1 ne s'applique qu'à la croyance partielle et elle ne s'applique pas à la croyance totale car, comme le dit Ramsey, α pour sûr est équivalent à α si p, β si $\neg$ p. En effet, en utilisant une propriété de la négation (que Ramsey peut maintenant démontrer), $Cr^{\circ}(p) = 1 - Cr^{\circ}(\neg p)$, comme α est sûr on note que

$Cr°$ ($\neg$ p) = 0, si bien qu'en utilisant la définition 8.2, la valeur espérée du pari $= (\alpha \times 1) - (\beta \times 0) = \alpha$. Si je suis certain de *p*, je n'ai pas de choix à faire, je n'ai qu'à attendre pour récolter α. Ramsey formule également un concept de degré de croyance relatif qui correspond à l'idée d'un pari conditionnel. Le degré de croyance en *p* étant donné *q* se représente par la cote à laquelle le parieur accepterait (maintenant) de parier sur *p* avec la condition que le pari ne tient que si *q* est vrai. Il est clair qu'un pari sur une conditionnelle n'a pas les mêmes conditions de succès et de satisfaction car un tel pari vaut dès maintenant. Il n'est pas non plus une explication adéquate du concept fort important de degré de croyance conditionnel que nous avons examiné à la section 3 du présent chapitre sous le nom de « *conditionnelle de Ramsey* ». La différence entre le degré de croyance en *p* étant donné *q* et l'idée intuitive de degré de croyance qu'on accorderait à *p* si on savait que *q* est que le concept de croyance conditionnelle qui est défini ici suppose l'indépendance de *q* et de *p*. Or, le fait de savoir que *q* peut altérer substantiellement mon évaluation de *p* change la donne. Ainsi, il n'est pas équivalent pour une personne en début de carrière de choisir l'une ou l'autre des options suivantes :

p : devenir fonctionnaire ;

p si *q* : devenir fonctionnaire si la situation économique s'améliore ;

p si $\neg$ *q* : devenir fonctionnaire si la situation économique ne s'améliore pas.

Il est clair que ma connaissance de la situation économique aura une influence sur la valeur que j'accorde à la sécurité que procure un emploi de fonctionnaire. Ramsey mentionne cette difficulté et il conçoit clairement que le concept de degré de croyance relatif qu'il propose est une simplification. Sur le plan logico-philosophique, cet exemple constitue un bon argument pour refuser l'atomisme logique et reconnaître que la possibilité pour chaque

proposition d'être vraie ou fausse indépendamment de toutes les autres est une simplification formellement utile mais philosophiquement déraisonnable. Il faut remarquer que cette difficulté, le présupposé d'indépendance des propositions élémentaires, affecte également les théories de Bruno de Finetti et Savage. Dans l'optique de la présente thèse, la difficulté que nous venons de mettre en évidence indique la nécessité de rendre plus explicite l'infrastructure logique de la logique de la décision, et en particulier la logique de l'action qui est habituellement peu détaillée dans les théories axiomatisées.

définition 2.8.3 : le degré de croyance en *p* si *q*, en abrégé, *Cr*° (p | q)
Soit un agent pour lequel les deux options suivantes sont équivalentes :

	option 1	option 2
q	α	
¬ *q*	β	β
p & *q*		γ
¬ *p* & *q*		δ

Le degré de croyance en *p* si *q* est donné par la formule suivante

$$Cr^{\circ}(p \mid q) =_{df} \frac{\alpha - \delta}{\gamma - \delta}$$

On peut comprendre le sens de cette définition (2.8.3) en inspectant la fonction à partir d'un exemple. Soit un tournoi éliminatoire de football où la France doit battre l'Italie pour pouvoir affronter le Brésil.

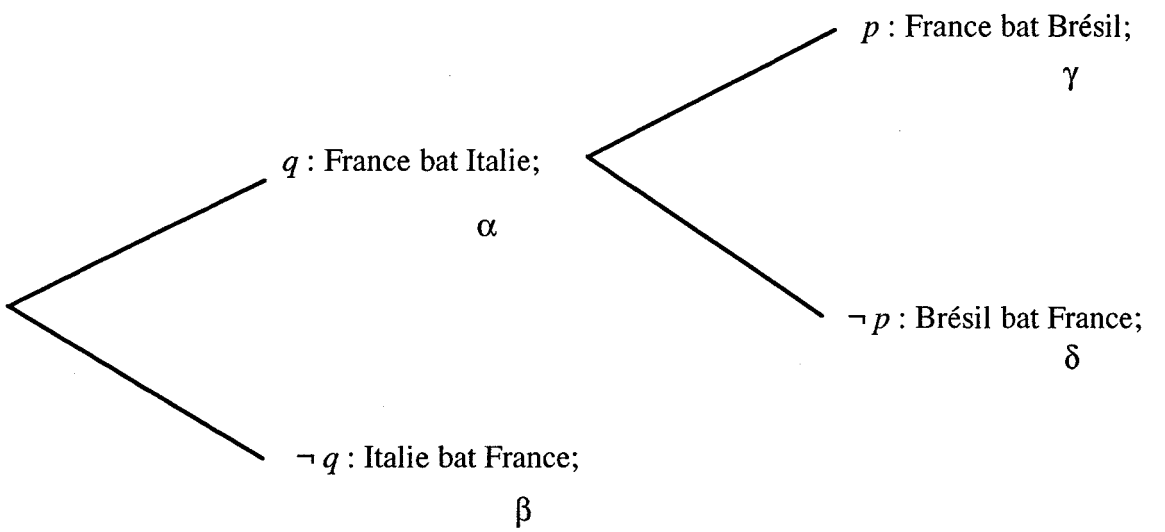


figure 2.8.1 le degré de croyance relative en p si q.

La figure 2.8.1 représente la situation où l'agent accepte de parier sur la victoire de la France contre le Brésil si la France parvient d'abord à vaincre l'Italie. Si la France perdait contre l'Italie, l'agent ne s'engage à aucun pari. Par comparaison avec la définition 2.5.1, on note que c'est ici la victoire de la France contre le Brésil q qui constitue le point de référence α, comme condition nécessaire du pari. Dans le calcul du degré de croyance relative, γ représente l'avantage de gagner et δ l'inconvénient de perdre le pari sur p.
Comme il se doit, la valeur β associée à la négation de l'antécédent de la conditionnelle, $\neg q$, ne joue aucun rôle dans le calcul de la croyance conditionnelle $Cr^{\circ}(p \mid q)$.

À partir des définitions et des axiomes qui précèdent, Ramsey annonce qu'on peut prouver les lois fondamentales de la croyance probable, chacune étant une loi du calcul des probabilités. Nous allons exprimer ces lois dans la notation déjà introduite. Ainsi, les propositions qui suivent vont demeurer vraies si on remplace $Cr°$ par Pr.

(1) $Cr°\ (p) + Cr°(\neg p) = 1$

(2) $Cr°\ (p \mid q) + Cr°\ (\neg p \mid q) = 1$

(3) $Cr°\ (p \,\&\, q) = Cr°\ (p) \times Cr°\ (q \mid p)$

(4) $Cr°\ (p \,\&\, q) + Cr°\ (p \,\&\, \neg q) = Cr°\ (p)$

À propos de l'inter-dérivabilité de ces postulats, on constate que (1) et (2) sont des conséquences immédiates des définitions. Ramsey donne une preuve de (3) et montre que (4) se déduit facilement de (2) et (3)[89]. Il faudrait ajouter à cette liste une loi pour la disjonction, que Ramsey ne mentionne pas, et qui ajoutée aux autres, donne ce qui manque pour montrer que $Cr°$ est une mesure de probabilité[90]. On verra cette loi dans les deux formes qu'elle peut prendre :

loi de disjonction spéciale :

Si p et q sont des propositions incompatibles (c.-à-d., mutuellement exclusives),

(5) $Cr°\ (p \vee q) = Cr°\ (p) + Cr°\ (q)$

Preuve:

de (4), en substituant $(p \vee q)$ à chaque occurrence de p dans (4) on obtient

$Cr°\ (p \vee q) = Cr°\ ((p \vee q) \,\&\, p) + Cr°\ ((p \vee q) \,\&\, \neg p)$.

La preuve procède en réduisant chacun des termes de l'addition qui forme le membre de droite de cette égalité. Puisque $((p \vee q) \,\&\, p) \equiv p$ est une tautologie,

on acceptera l'égalité $Cr^{\circ}(p \vee q) \& p = Cr^{\circ}(p)$. Ce qui permet la réduction du premier terme :

$Cr^{\circ}(p \vee q) = Cr^{\circ}(p) + Cr^{\circ}((p \vee q) \& \neg p)$

Pour le second, on note que puisque p et q sont incompatibles, $q \supset \neg p$;

or de $q \supset \neg p$ on peut déduire que $((p \vee q) \& \neg p) \equiv q$. Cette équivalence permet la réduction du second terme, d'où l'on obtient

(5) $Cr^{\circ}(p \vee q) = Cr^{\circ}(p) + Cr^{\circ}(q)$.

Pour *Pr,* on peut prouver une version généralisée de (5) (*l'additivité générale finie*) pour une disjonction à n termes[91].

Nous avons utilisé deux instances d'un principe qui est manifestement trop fort dans sa formulation générale, à savoir que pour A, B des propositions simples ou complexes,

si $\vDash A \equiv B$, alors $Cr^{\circ}(A) = Cr^{\circ}(B)$

Ce principe est clairement valide pour *Pr* mais il peut paraître trop fort pour Cr°. À la suite de Jeffrey, on pourrait l'accepter en y voyant une conséquence inévitable du fait de prendre les propositions comme objets de la croyance[92]. Notre position est de ne pas accepter le principe dans sa formulation générale mais d'en accepter certaines instances, ici, la substitution des équivalences élémentaires indiquées. Par ailleurs, la difficulté qui se pose dans ce contexte illustre et vient soutenir notre conviction qu'il est nécessaire de recourir à une théorie des propositions pour laquelle le critère d'identité des propositions est plus fin que celui de la conception classique (c.-à-d., l'identité des conditions de vérité). Remarquons en passant, que la tentative de prouver (5) sans utiliser les substitutions logiques indiquées nous entraîne inévitablement dans des dérivations utilisant des lois arithmétiques — opérations algébriques sur les

fractions — pour lesquelles on pourrait également soulever le problème de savoir si elles sont cognitivement réalisées pour un agent quelconque. Cette remarque s'applique en particulier à la discussion de Sahlin dans Sahlin [1990] qui utilise de telles transformations (en omettant plusieurs étapes) pour établir (5).

loi de disjonction générale : (sans restriction sur p et q)

(5') $Cr^{\circ}(p \vee q) = (Cr^{\circ}(p) + Cr^{\circ}(q)) - Cr^{\circ}(p \,\&\, q)$

Cette loi est une généralisation de (5) obtenue en éliminant la restriction que p et q soient incompatibles et en remplaçant cette clause par un terme additionnel qui indique l'effet de considérer la possibilité de $(p \,\&\, q)$ sans « compter deux fois » les valeurs de p et de q .

2.8 Remarques sur l'adéquation de la théorie

Le système d'axiomes et de définitions pour Cr° que construit Ramsey n'est pas différent de la méthode des paris mais il est destiné à le remplacer à titre de fondement conceptuel. Comme il n'y a pas vraiment d'économie conceptuelle, on ne peut pas dire qu'il s'agit d'une réduction de l'ontologie. Cependant, et c'est ce qui importe pour Ramsey, on peut dire que les lois fondamentales des probabilités ont reçu une interprétation complète en termes de degrés de croyances et que la conception subjective des probabilités reçoit ainsi ses assises. Ramsey exprime de la façon suivante l'idée qu'il se fait de son explication

> On réalise, par conséquent, qu'une analyse de la nature de la croyance partielle révèle que les lois des probabilités sont

> des lois de consistance, une extension au domaine des croyances partielles de la logique formelle, la logique de la consistance. Leur signification ne dépend en rien de ce que quelque degré de croyance en une proposition soit précisément déterminé comme celui qui serait rationnel ; ils distinguent simplement certains ensembles de croyances qui respectent ces lois comme étant ceux qui sont consistants.[93]

Cette citation est de première importance pour comprendre le statut de la théorie qui vient d'être exposée. Le concept central est celui de consistance et pour Ramsey, il est la signature d'une théorie « formelle » au sens propre et littéral du terme, c'est-à-dire d'une théorie libérée de toute référence à des contenus psychologiques particuliers, aux performances ou aux limitations particulières des esprits des agents. On peut avancer plusieurs arguments pour soutenir cette conception de la logique, les meilleurs ayant probablement été proposés par Frege dans plusieurs écrits. Examinons ce que tout ceci implique pour l'ambition et les limites de ce type de projet. En un sens, ces éléments de réflexion sur le statut de la théorie se généralisent et s'appliquent à la logique de la décision en général et ce qu'en dira Jeffrey ne sera guère différent[94]. En montrant comment la théorie esquissée par Ramsey se donne comme un modèle qui sous-détermine son objet — la conformation de l'esprit d'un agent doxastique et ses contraintes de rationalité — nous allons étayer davantage notre parti pris pour une interprétation normative plutôt que descriptive de la logique de la décision.

Examinons d'abord le concept de consistance dans le présent contexte et sa relation au concept voisin de cohérence. Dans la délibération, il faut sans doute tenir compte de la cohérence dynamique et de la persistance des choix ;

l'infrastructure logique exposée au chapitre VI permet un traitement plus constructif et détaillé de cette question.

Ce que la logique classique, plus précisément le calcul propositionnel, nous apprend de la consistance d'un ensemble de propositions est peut-être trop bien connu pour qu'il soit nécessaire de le rappeler ici ; je l'indique rapidement pour fixer les idées. Étant donné le schéma de preuve par réduction à l'absurde la théorie de la déduction nous donne que

$$\text{si } \Gamma \vdash A \,\&\, \neg A, \text{ alors } \Gamma \vdash B$$

où Γ est un ensemble quelconque de formules. En d'autres termes, toute proposition B est déductible d'un ensemble inconsistant de prémisses. La transposition de tout ceci dans un contexte épistémique, c'est-à-dire l'interprétation épistémique de la théorie élémentaire de la déduction est une affaire délicate.

Soit un agent (épistémique) pour lequel Ω s'interprète comme l'ensemble de ses croyances (c.-à-d., sa sphère de croyance) exprimée par un ensemble de propositions Γ fermé sous l'opération de déduction. Si cet ensemble Ω comporte une contradiction, alors on dira de cet agent qu'il croit (est disposé à affirmer) toute proposition B. Évidemment, tout ceci est une approximation bien trop simplifiée de ce que serait une base raisonnable pour une logique de la croyance. En effet, ce raisonnement comporte au moins deux présupposés problématiques. Le premier est que les propositions (au sens de Wittgenstein ou en un autre sens qu'il faudrait clarifier) sont les objets véritables de la croyance. Le second est que la sphère des croyances d'un agent est fermée sous l'opération de conséquence logique au sens de la théorie de la déduction. Nous avons déjà émis quelques réserves pour le premier de ces deux présupposés en indiquant qu'une alternative serait proposée au

chapitre VI. Il en va de même pour le second, qui est un aspect de l'omniscience logique et qui pose une exigence manifestement trop forte pour une intelligence humaine ou artificielle. C'est dire dans quelle mesure la logique formelle au sens restreint expliqué plus haut avait peu à offrir au chapitre d'une théorie de la consistance des croyances à l'époque de Ramsey. Incidemment, on pourrait s'étonner de ce que Ramsey considère la logique formelle comme une logique de la consistance plutôt que comme une théorie de la validité des inférences ou une science des vérités nécessaires. Mais on se souviendra que cette façon de voir s'appuie sur une longue tradition, en notant au passage qu'elle sert mieux son propos dans le présent contexte.

Ramsey voit dans la théorie qu'il a esquissée une extension du domaine de la logique formelle et il l'affirme clairement dans le passage que nous venons de citer. Le concept qui permet ce passage est celui de consistance et non pas comme on aurait pu l'imaginer, une certaine conception de la raison pratique au sens de Kant. Dans la théorie de Ramsey, il y a une double exigence de consistance. Il est requis de tout agent possédant des croyances dont le degré est déterminé qu'il soit disposé à parier à une même cote peu importe l'enjeu exprimé en termes de biens ultimes. Il est bien clair que l'idée de mesure implique fatalement l'idée d'une certaine stabilité des grandeurs mesurées, ce que semblent oublier plusieurs critiques de la théorie de l'utilité. C'est une condition nécessaire. La seconde exigence de consistance est beaucoup plus importante. Elle est reçue comme un critère central de validation de la conception subjective des probabilités — ou plus généralement d'une logique de la décision qui incorpore cette interprétation de la théorie des probabilités ; elle est connue sous le nom d'argument du « dutch book ». Ramsey énonce cet argument sans en donner une preuve ou une

illustration[95]. Il est clair qu'il faut prêter attention au contexte dans lequel l'argument est introduit pour éviter des confusions sur son sens et sa valeur. Ramsey indique qu'un agent dont la disposition d'esprit (*mental condition*) ne respecterait pas les lois (1) à (5)[96] verrait ses choix dépendre de la façon (*the precise form*) dont les options lui sont présentées, et en particulier

> On pourrait l'engager dans une série de paris à son désavantage (*have a book made against him*) par un preneur au livre habile de telle façon qu'il perde, quoi qu'il advienne.[97]

Alors que Ramsey utilise uniquement le terme de consistance, il est nécessaire de distinguer la *consistance,* au sens habituel qui est familier en logique et la *cohérence* en réservant le second terme pour la propriété qui vient d'être expliquée — la conformation d'un esprit qui respecte les lois de la probabilité. La relation entre les deux termes de cette distinction soulève quelques interrogations intéressantes. Il semble possible de montrer que tout ensemble *incohérent* — qui n'est pas cohérent au sens qui vient d'être expliqué — entraîne nécessairement une *inconsistance* au sens épistémique, c'est-à-dire une disposition à affirmer et à nier simultanément une même proposition. Cette proposition aurait la forme « que tel pari est avantageux », ou plutôt « que telle série de paris est avantageuse ». Bien sûr, cette implication n'est pas une équivalence — la réciproque n'est pas vraie — car la disposition à affirmer une proposition contradictoire est liée à un cas très particulier de configuration mentale qui implique une distribution de valeurs pour les croyances partielles. En tout état de cause, il semble que ceci ne soit pas qu'une simple question de terminologie. La relation entre la consistance et la cohérence n'est pas analysée par les principaux auteurs dont nous avons consulté les écrits ; ils utilisent uniquement le concept de cohérence, à la suite

de De Finetti[98]. Une fonction de probabilité personnelle, autrement dit une distribution de degrés de croyance partielle au sens de Ramsey est cohérente si elle satisfait les lois (axiomes) de la théorie des probabilités[99]. Il est possible que Ramsey considérait la cohérence comme une extension de l'idée de consistance, comme il qualifie sa « logique humaine » d'extension de la logique formelle. Reste que la relation entre le concept de cohérence et le concept de consistance au sens élargi de l'interprétation épistémique demande à être expliquée. Nous proposerons un examen plus détaillé du problème qui se pose dans la logique sous-jacente à cette discussion ultérieurement.

On peut maintenant aborder la question du « dutch book »[100]. Ramsey énonce que si la conformation de l'esprit d'un agent ne respecte pas les postulats de la théorie des probabilités, on peut construire une série de paris qui devraient être acceptés par cet agent et qui sont tels qu'il va nécessairement perdre quoi qu'il advienne. De Finetti a démontré la proposition équivalente sous sa forme contraposée : Si un agent est cohérent, on ne peut construire un *dutch book* contre lui. Ceci lui permet d'affirmer que la cohérence est le principe unique sur lequel on peut fonder les lois de la probabilité[101]. La cohérence est une condition nécessaire pour éviter la série ruineuse des paris du *dutch book* mais est-elle une condition suffisante ? L'affirmer revient à soutenir que le respect des lois de la probabilité est la seule condition requise pour assurer l'immunité contre tout *dutch book* [102]. Ceci peut également être démontré ; c'est la converse du théorème du « dutch book ». Pour démontrer ce théorème, on montre que si votre fonction $Cr°$ satisfait les lois de la probabilité et que vous calculez bien la cote des paris qui vous sont proposés, ces paris auront des valeurs négatives et vous devriez les refuser. Remarquons au passage qu'il ne serait pas souhaitable de chercher à

réduire le concept de rationalité à celui de cohérence. La cohérence est une forme pure et minimale de rationalité, une condition nécessaire pour la rationalité dans une interprétation normative, mais qui n'est pas suffisante. Autrement dit, on attend d'une *théorie* de la décision qu'elle fournisse un concept plus riche. Nous devons renvoyer le lecteur qui douterait qu'un concept de rationalité plus substantiel puisse être formulé dans le cadre d'une logique de la décision — c.-à-d., selon une élaboration du calcul de l'utilité espérée — au calcul des « valeurs décisionnelles » (*decision value*) de R. Nozick[103].

On a dit que l'argument du *dutch book* constituait une sorte de validation des lois de probabilités dans la conception subjective. Ceci est particulièrement clair dans le cas de la converse du théorème du *dutch book.* Mais nous croyons que c'est aussi vrai dans la version qu'énonce Ramsey. Soulignons qu'il s'agit d'une proposition démontrée et que le concept de cohérence qui y figure en position centrale reçoit une interprétation pragmatique — en termes de paris ruineux. Ainsi, il est clair que ce concept a un contenu intuitif indéniable, même si l'interprétation vaut pour un agent idéalisé, un parieur dont le preneur au livre est Dieu tout-puissant (*God almighty*), etc.

Il reste à discuter un dernier aspect de l'adéquation formelle de la théorie de Ramsey qui concerne l'existence et l'unicité d'une fonction à valeurs réelles, définie sur tous les mondes, et qui est le « miroir » des différences égales de valeurs que postulent les axiomes. Le théorème qui affirme l'existence et l'unicité d'une telle fonction porte le nom de « théorème de représentation ». Il montre qu'une certaine structure, qui n'est pas intrinsèquement numérique, peut être *représentée* de façon numérique ; d'où

son nom. En logique de la décision et pour les axiomatisations qui sont des variantes ou des alternatives aux fondements proposés par Ramsey, un théorème de représentation a le statut de résultat d'adéquation qui valide une théorie, un peu comme une preuve de complétude est un résultat d'adéquation qui valide un calcul logique axiomatisé. On peut dire avec J. M. Joyce que la représentabilité probabiliste de l'ensemble des croyances partielles associées à un agent est le véritable critère bayésien minimal de rationalité[104].

Comme le note Sobel, il y a un peu de confusion autour de l'idée d'un théorème de représentation pour le système d'axiomes de Ramsey[105]. Après avoir énuméré ses axiomes Ramsey lui-même se limite à énoncer qu'ils « permettent de corréler les valeurs une à une avec des nombres réels [...] »[106] en représentant fidèlement les différences de valeurs. Ceci équivaut à affirmer l'existence de la fonction qui établit la corrélation ; Ramsey ne parle pas de l'unicité. Il ne démontre pas ce qu'il affirme dans le passage que nous venons de citer, pas plus qu'il n'indique comment cette proposition pourrait être démontrée ou qu'il aurait fait cette démonstration. La question est posée : est-ce qu'il y a un théorème de représentation chez Ramsey ? Comme le souligne Sobel, dans Sobel [1998], nous sommes forcés de constater que Ramsey ne nous a pas donné explicitement un tel théorème. Il est donc fort étonnant de lire dans Sahlin et Gärdenfors [1988], dans Jeffrey [1992]b, ou dans Joyce [1999], que Ramsey aurait bel et bien démontré l'existence et l'unicité de sa fonction qui est simultanément la fonction d'utilité et de probabilité[107]. Contredisant ce qui est clairement énoncé dans l'introduction rédigée avec Gärdenfors, on note que le même Sahlin, dans Sahlin [1990], affirme qu'il n'y a pas de preuve du théorème de représentation à trouver chez Ramsey et il esquisse la signification de ce théorème à partir d'un exemple. Il se peut que

cette confusion historique soit due au fait qu'on estime que Ramsey a réuni dans son système de définitions et d'axiomes tout ce qui semble nécessaire à l'élaboration du résultat qu'il annonce. Cependant, même cette idée — que l'on pourrait procéder à une déduction du théorème de représentation à partir des seules définitions et axiomes fournis par Ramsey — ne doit pas être considérée comme une évidence selon Sobel[108]. En tout état de cause, il semble que personne n'ait encore formulé un théorème de représentation qui soit propre et fidèle au système de Ramsey même si personne ne doute qu'un tel résultat soit démontrable. On sait que divers théorèmes de représentation existent pour des systèmes apparentés à celui de Ramsey[109].

Examinons maintenant les clauses du théorème de représentation à partir de la formulation de Sobel [1998] pour mieux saisir son importance. Pour tout agent dont les préférences ordonnent les mondes et certains paris sur ces mondes (et quasi-mondes) de façon à satisfaire les axiomes (exposés à la section 7),

(Existence) il existe une fonction à valeurs réelles V qui représente fidèlement les différences de valeurs entre ces mondes (comme nous l'avons indiqué au début de la section 8)[110] ;

(Unicité) une fonction à valeurs réelles représente fidèlement les différences de valeurs entre les mondes si et seulement si elle peut être obtenue par une transformation linéaire de V.

Pour comprendre l'importance de ces propositions, on peut poser la question de savoir ce que cela signifierait si elles n'étaient pas vérifiées. Pour l'existence, on note que si la fonction V telle que décrite n'existait pas, les probabilités subjectives des propositions, des valeurs et des mondes n'auraient

pas la structure métrique indiquée dans les axiomes et en ce sens, la théorie de Ramsey ne serait pas « fondée ». Pour l'unicité, la situation est moins claire. Si la fonction *V* n'était pas unique, la commensurabilité des différences de valeurs requise par l'axiome 5 ne serait pas vérifiée. Or, il n'est pas aisé d'interpréter l'axiome 5 comme un axiome de rationalité. C'est plutôt un axiome de structure. Ainsi, nous devons admettre qu'il n'est pas aisé de formuler intuitivement ce que l'unicité implique pour la hiérarchie de préférences de l'agent sans affirmer que l'existence de telles conséquences est inconcevable.

Ce genre d'interrogations indique la véritable motivation d'un théorème de représentation dans la théorie de Ramsey et dans les théories similaires. Comme le note Joyce, un théorème de représentation est

> la méthode standard pour justifier une version (quelconque) d'une théorie de l'utilité espérée [....] qui montre que l'agent dont les croyances et les préférences satisfont certaines contraintes formulées axiomatiquement se comporte automatiquement comme s'il maximisait l'utilité espérée telle que définie par cette théorie. Un tel théorème garantit que le concept d'utilité espérée de cette théorie est sensé (*makes sense*) et qu'il peut s'appliquer dans un large éventail de situations particulières.[111]

Il serait intéressant de montrer comment le théorème de représentation garantit que l'agent dont la structure de préférences respecte les axiomes se comporte comme s'il maximisait l'utilité espérée. Pour ce faire, il faudrait construire et déployer plusieurs préliminaires formels nécessaires à la construction d'un théorème de ce type ; comme nous l'avons déjà remarqué il semble que

personne n'a encore accompli complètement cette tâche pour la théorie de Ramsey[112].

2.9 Sommaire des caractéristiques et théories connexes

Comme on le disait au début de ce chapitre, la théorie de Ramsey n'est pas la seule axiomatisation de la conception subjective des probabilités. Si elle possède, *ex æquo* avec l'axiomatisation proposée par De Finetti, le privilège de l'antériorité historique, la théorie de Ramsey n'a pas été aussi influente que la théorie de l'utilité de J. von Neumann et O. Morgenstern ou l'axiomatisation de L. Savage. On sait aujourd'hui que ces théories sont formellement équivalentes[113]. En un sens qu'il nous appartient maintenant d'éclaircir, elle possède certaines propriétés qui lui appartiennent en propre et qui peuvent être considérées comme des avantages ou des défauts selon le cas. Fishburn a dressé un tableau assez complet des diverses axiomatisations comparables et plusieurs des observations de ce type sont énumérées dans Fishburn [1981]. D'autres éléments comparatifs sont décrits dans la littérature que nous avons déjà citée dans ce chapitre, dont Sahlin [1990], Savage [1954], Jeffrey [1965] et Sobel [1998].

Nous avons déjà commenté le fait que la théorie de Ramsey utilise le concept de proposition éthiquement neutre et présuppose l'existence de propositions atomiques en signalant que ces caractéristiques n'étaient pas essentielles à la théorie bayésienne et qu'elles pouvaient se comprendre comme des fictions qui correspondent à des idéalisations commodes servant à simplifier la théorie qu'envisageait Ramsey, une sorte de version schématique de ce qui se passe en réalité[114]. L'axiomatisation proposée par Davidson et

Suppes n'utilise pas les concepts de monde possible et de propositions et par conséquent, elle esquive les difficultés associées à ces concepts[115]. Comme nous l'avons indiqué précédemment, nous croyons qu'il est fort important de clarifier la logique sous-jacente à la logique de la décision et pour des raisons qui tiennent aux objectifs de la présente recherche, nous affirmons qu'on ne peut souscrire sans réserves et de façon *a priori* à l'idée de simplifier les distinctions entre les actes, les propositions, les conséquences, les mondes, etc. sans s'inquiéter des conséquences de cet appauvrissement pour la capacité expressive du langage dans lequel on cherche à exprimer des extensions conservatrices de ce qui sert de base à la logique de la décision. On ne peut simplement pas souscrire à l'affirmation de Suppes

> Il y a une tradition classique d'ambiguïté en théorie des probabilités à propos de savoir si on parle de la probabilité des événements ou de celle des propositions et je ne crois pas qu'il soit important de choisir l'un ou l'autre.[116]

En regard des seuls axiomes de la théorie des probabilités, on peut soutenir qu'il n'y a pas de différence ; mais pour la logique de la décision qui est construite comme une extension de la théorie bayésienne, nous croyons qu'il y a des conséquences logico-philosophiques significatives associées à ce choix et aux autres choix de ce type. Nous croyons l'avoir montré en discutant la théorie de Ramsey et nous pourrons le montrer encore dans les prochains chapitres. Nous nous bornons pour l'instant à offrir un argument indirect et plus ou moins indépendant de ce que nous disions ailleurs de cette question.

Comme Sobel et Joyce l'ont bien montré, il peut être utile de reconsidérer les fondements d'axiomatiques plus anciennes pour exprimer de

nouvelles idées et explorer de nouvelles avenues. Ainsi, Sobel montre que les bases de la théorie de Ramsey sont neutres par rapport aux théories évidentielles et causales de la décision tandis que Joyce formule son théorème de représentation pour ce qu'il appelle la théorie conditionnelle de la décision en utilisant comme base la théorie de Savage plutôt que celle de Jeffrey[117]. Dans leurs travaux qui sont très récents et qui témoignent de préoccupations dont l'actualité est indiscutable, ces auteurs nous aident à mieux comprendre les relations entre les théories évidentielles et les théories causales de la décision en se réappropriant des formulations antérieures qui pourraient sembler dépassées et inactuelles. Voilà ce que l'on peut considérer comme un argument établissant indirectement l'intérêt de la logique sous-jacente aux théories de la décision.

En discutant la théorie de Ramsey dans *The Logic of Decision*, Jeffrey note que la différence principale entre sa propre théorie et celle de Ramsey consiste en ce que la théorie des préférences qu'il propose attribue les valeurs de probabilité et de désirabilité aux propositions, éliminant la disparité qui existe dans la théorie de Ramsey entre les situations possibles considérées — ce que suivant Sobel, nous désignons par le terme de « quasi-mondes » — et les conséquences des choix, les résultats des paris. En fait, la théorie de Jeffrey élimine toute référence aux paris comme « entités hybrides » composées de situations (*conditions gambled upon)* et de conséquences (*possible outcome of the gambles*)[118]. Ainsi la théorie de Jeffrey représente un progrès dans la direction d'une formulation plus proprement logique dans la mesure où elle montre en détail comment remplacer la référence à des paris par des opérations élémentaires sur les propositions à l'aide des connecteurs usuels.

Dans l'appendice de la seconde édition de leur ouvrage sur la théorie des jeux, von Neumann et Morgenstern ont axiomatisé un concept d'utilité qui, sur le plan formel, est équivalent à l'axiomatisation de la croyance partielle de Ramsey[119]. Cependant, il faut remarquer que von Neumann et Morgenstern ne discutent absolument pas la question de la nature des probabilités et qu'ils se concentrent entièrement sur la notion d'utilité et de préférence. Pour comprendre comment ceci est possible, il faut se rappeler que la théorie de Ramsey axiomatise simultanément l'utilité, c'est-à-dire les mondes comme valeurs (*world-values*) et la probabilité, c'est-à-dire la croyance partielle.

Nous allons conclure cette section en soulignant un aspect de la théorie de Ramsey que Fishburn signale comme un inconvénient et qui a certainement été compris comme une difficulté potentielle pour plusieurs théories comparables[120]. Le problème concerne la cardinalité des mondes qui résultent des actes. Les axiomes de Ramsey conduisent à envisager l'existence d'une infinité de conséquences (mondes comme valeurs) et ceux-ci permettent de faire des distinctions arbitrairement fines dans l'échelle des préférences. D'un point de vue descriptif, le problème est double. Cette infinité de valeurs et cette capacité de discriminer de façon arbitrairement fine entre eux sont deux traits qui posent des problèmes d'interprétation : elles ne peuvent décrire l'esprit d'aucun agent humain imaginable. Comme le note Fishburn, le fait que la cardinalité des mondes soit infinie constitue une restriction ; les axiomes ne seront pas satisfaits dans une structure d'interprétation dont le domaine est simplement arbitrairement grand. Qu'il nous soit permis de remarquer en passant que ces questions de cardinalité ne sont pas fréquemment soulevées dans la littérature de la logique de la décision, à tout le

moins chez les auteurs que nous citons dans la présente thèse. La raison en est peut-être que certaines questions relatives à la cardinalité posent des problèmes difficiles à propos desquels personne n'a d'idées concluantes. Il est bien connu que les efforts pour se défaire de ces restrictions donnent lieu à des bases axiomatiques moins simples et des postulats qui n'ont pas d'interprétation intuitives comme le note Fishburn[121]. Pour cette raison, il faut admettre que les alternatives finitistes comportent leurs propres problèmes d'interprétation et tout compte fait, on peut s'interroger sur l'intérêt de l'avantage obtenu.

La recherche d'une axiomatisation qui repose sur une base finitiste est sans doute la réponse la plus évidente à ce problème de cardinalité ; mais ce n'est pas la seule option pouvant constituer une alternative. À la suite des travaux d'Abraham Robinson, on sait que le calcul des infinitésimaux a été réhabilité dans une théorie mathématique qu'on appelle « l'analyse non-standard ». La possibilité d'interpréter les axiomes de Ramsey avec les nombres hyperréels (qui sont des nombres réels non-standards) est évoquée au passage par Sobel. Dans la dernière décennie, les recherches mathématiques sur les infinitésimaux indiquent une autre voie qui semble fort prometteuse. John Bell a montré qu'on peut développer une théorie de la continuité dans le contexte de l'analyse infinitésimale lisse (*smooth infinitesimal analysis*) qui s'inscrit, comme l'arithmétique de Robinson, dans une approche constructiviste — c.-à-d., incompatible avec la logique classique, mais compatible avec la logique intuitionniste. L'analyse infinitésimale lisse semble en mesure de rendre compte de nos intuitions concernant la continuité de l'espace perceptuel, c'est-à-dire l'espace des géomètres plutôt que celui des analystes. Le calcul des différences infiniment petites, c'est-à-dire, des

infinitésimaux, semble vraiment offrir une approche naturelle et intuitive pour l'explication mathématique de nos estimations qui impliquent de telles différences. Il n'est pas déraisonnable d'espérer que ces recherches, qui appartiennent au domaine des fondements des mathématiques, vont apporter un nouvel éclairage et permettre de recadrer le problème d'interprétation posé par les fonctions à valeurs réelles. Nous mentionnons ces recherches pour souligner l'existence de perspectives qui sont de nature à calmer les inquiétudes soulevées par le problème de cardinalité[122].

Dans Fishburn [1981] où le problème de cardinalité est formulé, il n'est pas dit clairement que Fishburn considère lui-même ce trait de la théorie de Ramsey et d'autres théories similaires comme posant une réelle difficulté d'interprétation importante ou s'il y voit une contrainte dont certains ont voulu s'affranchir. Dans ce texte, son objectif est simplement d'exposer et de comparer une trentaine de théories normatives axiomatisant des conceptions subjectives de l'utilité. Limitons-nous à exprimer un avis sur la question dans le but d'atténuer l'impression qu'un problème insurmontable se camoufle derrière ce type de difficultés.

Nous avons abordé ce problème dans l'esprit de ce que nous avions déjà avancé en invoquant et en plaidant pour une interprétation normative de la théorie de Ramsey. Nous croyons que ceci constitue une solution provisoire. En clair, nous ne voyons pas de problème d'interprétation à ce que l'univers de mesure des croyances soit beaucoup plus riche que ce qui pourrait être envisagé par un agent humain dans une situation de délibération réelle. On cite fréquemment l'étude de G. A. Miller qui fixe à sept, le nombre maximum de morceaux (*chunks*) d'information que l'esprit humain pourrait traiter dans le

champs d'attention actif de la mémoire[123]. En dernière analyse, il faudra bien tenir compte des limitations de cette nature.

En examinant la question de plus près, on note que même Suppes, qui a proposé une approche finitiste, admet que les théories de Ramsey et de Savage sont acceptables si on se limite à une interprétation normative[124]. On peut penser que Ramsey nous a donné une indication de la façon dont il envisage cette question à propos de problèmes de décision dans lesquels le nombre d'options seraient infini. Ce n'est pas exactement le même problème, mais c'est un problème de même nature. Vers la fin de son article, en faisant le bilan de ce qui a été accompli, il écrit :

> Troisièmement, rien n'a été dit à propos du degré de croyance lorsque le nombre d'options est infini. À ce sujet, je n'ai rien à dire d'utile sinon que j'ai des doutes sur la possibilité que l'esprit puisse considérer plus qu'un nombre fini d'options. Il peut concevoir des questions pour lesquelles une infinité de réponses sont possibles mais pour envisager les réponses, il doit les rassembler pour former des groupes en nombre fini[125].

Il me semble que la même orientation peut être adoptée en regard du problème de cardinalité soulevé par Fishburn. Voici comment nous croyons que la difficulté doit être contournée. C'est la distinction entre la cardinalité du référentiel de la théorie axiomatisée et la cardinalité des classes d'équivalences qui se présentent comme des conséquences qu'un agent réel pourrait ordonner par une relation de préférence dans un problème particulier de décision qui permet d'éviter la difficulté. Autrement dit, en appliquant la théorie à un problème particulier, on n'utilise pas toute la « richesse » de l'ontologie de la théorie. Dans ce contexte, le recours aux cardinalités infinies

n'est qu'une approximation ou une simplification de ce qui se passe en réalité et qui est certainement d'ordre fini. Cette façon de contourner le problème est dans l'esprit de l'orientation méthodologique que formule clairement Jeffrey :

> Pour le probabilisme radical, la question importante à notre sujet n'est pas de savoir si nous avons des probabilités dans notre esprit, mais si nous pouvons fabriquer des constructions de remplacement (*proxies*) pour quoi que ce soit qui pourrait se trouver dans notre esprit.[126]

Cette orientation implique que la logique sous-jacente dans laquelle les constructions de remplacement sont construites soit plus riche et permette de faire des distinctions plus fines que celles qui seraient admissibles pour une théorie dont l'interprétation visée est de nature descriptive.

Ramsey a montré comment on pouvait extraire les probabilités (degrés de croyance) et les degrés de préférence à partir de la façon dont l'agent ordonne ses préférences pour certains paris. Ce faisant, il parvient à formuler des contraintes de consistance qui valent pour tous les agents possibles et donne des fondements d'une logique de la décision. Comme le remarque Sahlin, le fait que l'essai de Ramsey soit centré sur la représentation des croyances partielles par des probabilités ne doit pas nous faire oublier qu'il comporte les fondements complets d'une logique de la décision. On y trouve une théorie de l'utilité, essentiellement équivalente à celle de von Neumann et Morgenstern, la réduction de la valeur d'une option dans une alternative (dans un problème de décision) à cette seule valeur d'utilité, l'assignation de valeur numérique aux conséquences possibles des choix, le postulat implicite de l'indépendance entre les actes et les états et la règle de décision fondamentale

dérivée du concept d'espérance mathématique qui recommande de maximiser l'utilité espérée.

Sur le plan philosophique et méthodologique l'apport de Ramsey est fondamental à plus d'un titre. Il a conçu la théorie de la croyance partielle comme une extension de la logique et l'a développée en s'efforçant de ne pas faire basculer son projet dans une théorie psychologique de l'inférence probable. De plus, il a édifié les fondements de sa théorie sur une explication qui est d'abord conceptuelle ou qualitative et qui ne fait intervenir la quantification que d'une façon dérivée à partir de l'explication en termes de paris. Il a reconnu le caractère essentiellement normatif du concept de choix rationnel. Avant Carnap, Ramsey a proposé de distinguer entre deux concepts de probabilité sans rejeter la conception fréquentielle. Enfin, il nous a tracé une voie qui permet d'éviter ou de contourner les difficultés liées à la problématique traditionnelle de l'inférence inductive en donnant des bases purement doxastiques — plutôt qu'épistémiques — à l'analyse de l'inférence probable.

CHAPITRE III

L'EXPLICATION DU CHOIX RATIONNEL CHEZ CARNAP

3.1 Introduction[1]

À première vue, la formulation d'une règle pour le choix rationnel apparaît dans les écrits de Carnap comme une étape plus ou moins essentielle dans l'édification de la logique inductive. Notre objectif premier est de montrer qu'il existe une définition opératoire partielle du concept de choix rationnel chez Carnap et d'en cerner les caractéristiques en suivant son évolution. De son propre aveu, Carnap présente une règle qui est analogue à celle qui est proposée habituellement (maximiser l'utilité espérée) et son originalité principale est de chercher d'abord à l'interpréter dans le cadre de la conception dite « logique » des probabilités[2]. Nous allons examiner l'évolution de l'explication du choix rationnel chez Carnap depuis les sections § 50-51 de l'ouvrage de 1950 *Logical Foundations of Probability*, la version de 1960 dans « The Aim of Inductive Logic »[3], celle de 1963 « My basic conceptions of probability » jusqu'à celle présentée dans les premières sections de l'ouvrage de 1971, *Studies in Inductive Logic and Probability,* qui est une version remaniée de celle de 1960. Le projet d'une logique inductive telle que la concevait Carnap est généralement déconsidéré aujourd'hui[4]. Le problème principal invoqué par les critiques est que les systèmes de logique inductive qui furent construits par Carnap requièrent des mesures de

probabilité *a priori* qui semblent forcément arbitraires et donc philosophiquement inacceptables. On pourrait répondre à cette objection qu'en faisant l'hypothèse additionnelle (relevant d'un empirisme de type bayésien) qui veut que l'expérience vienne éventuellement nettoyer les probabilités fausses ou arbitraires. Quoi qu'il en soit de la valeur de cette réponse et des mérites de la logique inductive de Carnap, nous aurons peu de choses à dire à sur l'épistémologie de Carnap[5]. Comme on le verra, certaines difficultés inhérentes à l'approche de Carnap affectent également son explication du choix rationnel. Cependant, nous croyons que l'idée de considérer les postulats de la théorie de la décision comme des contraintes qui donnent une interprétation précise et substantielle du concept de rationalité est une idée intéressante sur le plan méthodologique et qu'elle est bien fondée. L'approche de Carnap est parfaitement sensée du point de vue du courant probabiliste, courant dont l'expansion récente est notoire dans les domaines de la logique philosophique et de l'épistémologie. Notre hypothèse de départ en relisant Carnap était qu'il est possible de dégager chez lui une contribution significative au programme d'une théorie normative de la décision de type bayésienne. De plus, il nous semblait que le traitement du choix rationnel par Carnap pose bien l'objet et la méthodologie de l'ensemble de ce programme de recherche. Si on garde à l'esprit le fait que plus de 25 ans de recherches intensives nous séparent de la publication du tome 1 des *Studies in Inductive Logic and Probability*, on peut pratiquer une lecture charitable de Carnap. De 1950 à 1971, l'évolution de sa conception met en évidence des enjeux qui sont encore actuels pour la logique du choix rationnel et elle témoigne de progrès importants dans l'analyse philosophique de la rationalité pratique après 1950. De façon générale, on sait que l'élaboration théorique de Carnap et surtout sa

conception des probabilités se situe d'abord dans la lignée de J.M. Keynes [1921], H. Jeffreys [1939], B. de Finetti [1937] pour ensuite se rapprocher de celle de L. J. Savage [1954].

2. La première formulation (*circa* 1950)

Dans *Logical Foundations of Probability*, § 50 et § 51, Carnap introduit la question des décisions pratiques comme l'une des deux applications utiles de la logique inductive qu'il veut construire, l'autre étant la clarification des fondements de l'induction[6]. Une longue tradition associe l'utilité pratique du calcul des probabilités à l'impossibilité de connaître le futur avec certitude. Carnap s'inscrit volontiers dans cette tradition et il peut reprendre à son compte la formule célèbre de l'évêque Joseph Butler, « *probability is the very guide of life* »[7], qui place la probabilité au centre des délibérations qui engendrent nos choix. Au XIX^{e} siècle, des mathématiciens comme S.-D. Poisson et A. A. Cournot proposaient de distinguer la *probabilité* de la *chance*. Pour la philosophie contemporaine, c'est sans doute Carnap qui sera le défenseur le plus reconnu de cette distinction. Dans la terminologie de Carnap, la probabilité proprement dite est la probabilité logique ou probabilité$_1$. L'autre concept, qui correspond à l'idée de chance, s'explique par la fréquence observée. C'est le concept de probabilité$_2$. Pour Carnap, les deux concepts sont irréductibles l'un à l'autre et également légitimes; le problème étant de formuler une règle du choix rationnel dans une situation d'incertitude, Carnap y voit une occasion de plaider en faveur du concept de probabilité$_1$ (probabilité logique). En effet, lorsqu'une décision pratique dépend de la connaissance d'une certaine grandeur, par exemple de la fréquence avec

laquelle on retrouve une certaine propriété dans une population, cette décision doit se fonder sur une approximation (*estimate*) de cette grandeur relativement à certaines observations empiriques et non sur une connaissance de la fréquence réelle avec laquelle cette propriété se retrouve dans la population. De plus, il s'agit bien d'un concept de probabilité *logique;* un certain ensemble d'observations empiriques étant fixé[8], l'énoncé d'une probabilité relative à cet ensemble exprime bien une proposition analytique. Considérons par exemple l'énoncé suivant:

(1) L'approximation de la fréquence relative de la propriété *M* dans la population totale par rapport à la base empirique *e* est de 0,73.

Il est clair que les conditions de vérité de (1) découlent entièrement de la définition de l'approximation d'une fonction et de la base empirique *e* et l'énoncé lui-même ne peut être ni confirmé ni réfuté par des observations futures[9]. On voit donc se dessiner le modèle proposé par Carnap pour le choix rationnel. Pour effectuer une décision pratique, on part d'une base de connaissances empiriques qui peut contenir divers énoncés observationnels et on utilise la logique inductive fondée sur le concept de probabilité$_1$ pour attacher des degrés de confirmation aux hypothèses envisagées. C'est une particularité de cette première version du traitement que fait Carnap du choix rationnel que de présupposer un agent qui a en sa possession un système de logique inductive sous la forme d'une fonction de probabilité$_1$. Dans certains textes ultérieurs comme Carnap [1962]a et Carnap [1971], le point de départ de la construction du concept de choix rationnel ne présupposera plus que l'agent soit en possession d'une logique inductive. On supposera plutôt que l'agent possède une fonction de probabilité subjective comme chez Savage,

autrement dit une fonction qui assigne des degrés de croyance. Mais, dans *Logical Foundations*, on suppose que l'agent a accès à toutes les observations antérieures qu'il a effectuées et qu'elles sont compilées dans un rapport qui les énumère. Cet ensemble forme sa base empirique *e*. De plus, on suppose que l'agent est en possession d'une règle qui lui permet de calculer, sur la base de *e*, un degré de probabilité pour chaque hypothèse *h* qui peut l'intéresser. C'est en cela que consiste le fait d'avoir « une logique inductive » ou « un concept de probabilité$_1$ » pour un agent. Si on ajoute les concepts psychologiques de valeur et de préférence, on peut formuler une règle du choix rationnel pour un tel agent. Carnap souligne toujours avec insistance que les difficultés de ce programme sont entièrement liées à ces concepts de la psychologie et non à la logique inductive elle-même.

C'est par une discussion critique de cinq règles de décision que Carnap élabore progressivement l'explication du choix rationnel. La discussion critique des premières règles conduit progressivement vers la formulation d'une règle assez générale pour ne pas se heurter à des contre-exemples apparents.

Règle FR$_1$ (Règle de la probabilité élevée) Supposez que les événements qui ont une haute probabilité$_1$ relativement à la base empirique *e* vont se produire et agissez comme s'ils allaient effectivement se produire.

Cette règle ne donne pas toujours le choix optimal et elle ne s'applique pas lorsque nul choix ne possède une probabilité élevée. On peut aisément le constater à l'aide d'un exemple, comme nous le verrons après avoir considéré la seconde règle. Pour la formuler, nous avons besoin du concept de *partition*

d'un ensemble d'événements. On appelle partition d'un ensemble *A*, une famille non-vide de sous-ensembles non-vides et mutuellement disjoints de *A*, dont la réunion est *A*. Dans la terminologie utilisée par Carnap, il s'agit d'un ensemble exhaustif d'hypothèses mutuellement L-disjointes et mutuellement L-exclusives relativement à *e*. Notons au passage que dans le système formel de Carnap [1950-4], les probabilités sont distribuées sur des énoncés plutôt que sur des événements[10].

Règle FR_2 (Règle de la probabilité maximale) Relativement à une partition d'événements, prévoyez que l'événement possédant la probabilité$_1$ la plus élevée se produira et faites comme si vous saviez que cet événement se produira certainement.

Cette règle est mise en difficulté par l'exemple qui suit. Supposons qu'un restaurateur doive décider à l'avance combien de personnes feront le choix du plat Z dans le menu. Les probabilités sont distribuées de la façon suivante: Pour *n* clients où *n* prend successivement les valeurs « 0, 1, 2, 3, 4, 5, 6 », les probabilités sont respectivement « 0.20, 0.19, 0.18, 0.17, 0.16, 0.10, 0 ». Comme la situation la plus probable est que personne ne choisira le plat Z, la règle FR_2 recommande de ne pas le préparer. Or, cette éventualité n'a qu'une probabilité de ⅕ alors que la probabilité qu'au moins une personne demande le plat Z a une probabilité de ⅘. Ce sont les deux règles, FR_1 et FR_2, qui sont de mauvaises conseillères dans un tel cas.

Règle FR_3 (Règle de l'usage d'approximations) Supposons que votre décision dépende d'une certaine grandeur *u* que vous ne connaissez pas, au sens où, si

vous connaissiez *u*, alors cette connaissance de la valeur *u* déterminerait votre décision (autrement dit, il y a une fonction *F* telle qu'une propriété de votre décision prendrait la forme *F(u)).* Dans ce cas, calculez l'approximation *u'* de la valeur *u* relativement à l'ensemble de connaissances empiriques *e* et agissez à certains égards comme si vous saviez avec certitude que la valeur de *u* est *u'* (c.-à-d., posez que la propriété en question prend la valeur *F(u')*) ou une valeur voisine de *u'*.

On peut montrer que cette règle est supérieure aux deux règles précédentes. En particulier dans le cas du restaurateur, l'approximation du nombre de personnes qui demanderont le plat Z étant de 2.2. Cette valeur est la somme des produits

$$(0 \times .2) + (1 \times .19) + (2 \times .18), \ldots, + (6 \times 0)$$

où les valeurs sont tirées des deux suites données plus haut. Ce calcul donne une estimation (*probability$_1$-mean estimate*) au sens de Carnap [1950][11]. Ainsi, le restaurateur va prévoir et préparer deux fois le plat Z. L'expression « à certains égards » est vague et indique une des faiblesses de cette règle. Si le restaurateur agissait à tous égards comme si la valeur *u* était *u'*, il serait disposé à parier à mille contre un que deux clients seulement demanderont le plat Z, ce qui est déraisonnable. La règle FR_3 est également désavantageuse pour le décideur dans les situations où les pertes encourues ne sont pas distribuées symétriquement selon qu'il y a surestimation ou sous-estimation de la valeur de *u*. Pour une cuisine donnée, il pourrait y avoir une plus grande perte financière à sous-estimer le nombre de clients d'une certaine quantité qu'à le surestimer de la même quantité.

Règle FR_4 (Maximisation du gain espéré par approximation)[12] Parmi les actions possibles, choisissez l'action pour laquelle l'approximation du gain, déterminée à l'aide des probabilités assignées aux divers résultats possibles, n'est pas inférieure à celle de tout autre action. Si plusieurs actions ont pour résultat l'obtention de la valeur maximale de l'approximation, vous pouvez choisir l'une ou l'autre de ces actions car il est indifférent de choisir une action plutôt que l'autre.

Carnap propose aussi une version de la règle FR_4 en termes d'une offre que l'on peut accepter ou refuser. Cette version, **FR_4*** repose sur les définitions évidentes des concepts « d'offre favorable », « d'offre défavorable » et « d'offre équitable » en termes de la définition du gain espéré par approximation.

Règle FR_4*. Si on vous fait une offre qui est favorable, acceptez cette offre; si elle est défavorable, refusez l'offre; si l'offre est équitable, vous pouvez l'accepter ou la refuser.

Les règles FR_4 et FR_4* ont le mérite de ne pas exiger qu'on agisse comme si on savait réellement ce qu'on ne sait pas. Encore une fois, elles sont supérieures aux règles précédentes, mais elle ne sont pas parfaites. On appelle *utilité marginale* d'un bien pour quelqu'un, la quantité d'unités de valeur (efforts, énergie ou argent) que la personne est disposée à offrir en échange de ce bien. On peut montrer que les règles FR_4 et FR_4* conduisent à des décisions qui ne sont pas optimales dans si la fonction qui représente l'utilité marginale d'un bien en fonction de sa quantité n'est pas linéaire[13]. Il est bien

connu que c'est la raison pour laquelle l'argent n'est pas une bonne représentation de l'utilité au sens de « valeur subjective ». Ainsi, il peut être raisonnable pour moi de parier 1$ pour en gagner 10$, si le pari m'est légèrement favorable, mais il ne serait pas raisonnable pour moi de parier 10,000$ dollars pour risquer d'en gagner dix fois plus, 100,000$, car il n'est pas raisonnable du tout de risquer une si grande part de ma fortune. Dans ce cas, il est rationnel de refuser un pari qui m'avantage. Inversement, compte tenu des profits des compagnies d'assurances, il est clair que de façon générale, l'achat d'un contrat d'assurance peut être considéré comme un pari qui est défavorable à l'acheteur. Cependant, si la prime de l'assurance est relativement peu onéreuse, en tenant compte de la faible utilité marginale d'une petite somme, il peut sembler désirable de l'acheter. Ce faisant, j'accepte un pari qui m'est défavorable au sens de FR_4*.

Après avoir exposé le principe de la diminution de l'utilité marginale de Bernoulli (et Cramer) qui était bien illustré dans l'exemple précédent de la paire de paris favorables, Carnap formule le problème de la mesurabilité de l'utilité et se trouve en position d'énoncer la règle usuelle (maximisez l'utilité espérée) qui tire sa signification des efforts requis pour surmonter les inconvénients des quatre premières règles. Ainsi la généralité et la simplicité de la règle FR_5 ne doit pas nous la faire considérer comme une évidence.

Règle FR_5 Parmi les actions possibles, choisissez celle pour laquelle l'approximation de l'utilité du résultat est maximale.

Avec cette règle, l'objectif de proposer une explication de la règle du choix rationnel comme application de la logique inductive est atteint. Carnap ne dit pas explicitement que cette formulation d'une règle pour la décision est une *explication*, au sens spécial que Carnap associe à ce terme, du concept de décision rationnelle, mais on peut en juger facilement en vérifiant qu'une telle règle satisfait (ou devrait satisfaire) les quatre critères qu'il formulait au début de son ouvrage[14]. Il n'est pas difficile de constater que c'est bien le cas.

En premier lieu, la règle qui doit servir *d'explicatum* au choix rationnel est *similaire à l'explicandum.* Elle l'est pour autant qu'elle permet de faire les bonnes prédictions concernant les choix qui vont s'avérer avantageux pour le décideur. En second lieu, la règle du choix rationnel doit recevoir une *formulation exacte* reliée à un réseau de concepts scientifiques bien définis. Là encore, en principe, c'est bien ce qui est souhaité pour une règle du choix rationnel. Carnap exprime de l'insatisfaction à l'endroit du caractère vague de l'expression « à certains égards » dans FR_3. En troisième lieu, une explication doit être *fructueuse* et conduire à la formulation de plusieurs énoncés universels. En un sens, c'est bien ainsi que vont les choses car la règle du choix rationnel nous conduit à chercher une formulation précise et quantifiable de la notion d'utilité. Enfin, en regard de la dernière exigence, on peut soutenir que la règle FR_5, telle qu'elle est formulée et dans la forme quantifiée qu'elle recevra bientôt, satisfait le critère de *simplicité*. La définition du concept de valeur subjective d'un acte est simple du point de vue mathématique, c'est-à-dire dans l'univers des fonctions. Bien sûr, cette version du concept de simplicité n'est pas celle du pédagogue ou du néophyte.

Si on compare la première explication proposée par Carnap pour le concept de choix rationnel aux théories économiques de l'utilité comme celles de Samuelson ou de von Neumann et Morgenstern mentionnées dans Carnap [1950-4], on constate que la version de Carnap possède deux caractéristiques. D'une part il y a l'usage du concept de probabilité$_1$ (probabilité logique ou probabilité relative de *h* par rapport à *e*) dans la règle de décision et conjointement, il y a cette idée étendard du bayésianisme qui est d'utiliser l'approximation des valeurs de probabilité comme façon de tenir compte de notre ignorance relative du futur et de l'imprécision de notre connaissance des grandeurs mesurables[15].

Dans son autobiographie intellectuelle, Carnap indique clairement l'idée qu'il se fait de sa contribution:

> Sur la base indiquée [la fonction *c* du degré de confirmation], il est possible de formuler une règle qui détermine le choix rationnel pour une personne *X* parmi un ensemble de décisions possibles qui peuvent être formulées. [...] Cette règle est analogue à celle que l'on retrouve dans les conceptions habituelles. Mais il me semble que ma version de la règle est plus adéquate que les formulations habituelles car elle utilise le concept de probabilité logique et non le concept de probabilité statistique ou d'autres concepts statistiques. Il me semble clair qu'une règle du choix rationnel de *X* au temps *T* ne doit utiliser que la connaissance e qui est disponible pour *X* au temps *T*. Les valeurs pertinentes de probabilité statistique sont en général, toutefois, inconnues de *X*; par conséquent, la règle ne devrait pas y référer. Par ailleurs, les valeurs de probabilité logique sont déterminées par des procédures purement logiques sur la base d'observations données.[16]

3. Carnap [1962]a et la version augmentée Carnap [1971]

C'est dans la conférence intitulée « The Aim of Inductive Logic » prononcée au congrès de Stanford en 1960 que Carnap propose la seconde formulation de l'explication du choix rationnel. La formulation la plus élaborée de sa théorie, exposée dans Carnap [1971] « Inductive logic and rational decision », est en réalité une révision et une extension de la version 1962 [17]. Nous verrons que l'évolution vers le personnalisme bayésien est évidente dans son traitement de la question du choix rationnel. Alors qu'en 1950, à la suite de Keynes, il rejetait l'expression de « probabilité subjective » à cause de ses connotations psychologistes indésirables[18], il débute ici en disant

> Je vais essayer de montrer que nous devons comprendre « probabilité » dans ce contexte [la théorie de la décision], non pas au sens objectif, mais au sens subjectif, c.-à-d. en tant que degré de croyance. C'est un concept psychologique de la théorie de la décision empirique, qui réfère aux croyances réelles d'êtres humains réels.[19]

Dans la version de 1971, il utilisera plutôt l'expression « probabilité personnelle » qu'il reprend de Savage [1954] où elle est apparue[20]. Dans la théorie de Savage, les probabilités subjectives qui satisfont le critère de cohérence sont appelées « probabilités personnelles ». On trouve dans la version 1962 une conception structurée en trois étapes qui évolue vers la logique inductive proprement dite.

La première étape est la théorie descriptive ou théorie empirique de la décision qui réfère aux croyances réelles d'humains véritables. Il s'agit d'une théorie comportant des concepts empiriques où l'on traite de *décisions réelles*. Par

abstraction, on passe ensuite à la théorie normative de la décision qui explique le concept de *décision rationnelle*. En éliminant complètement le contenu quasi-psychologique du concept de choix rationnel, on arrive à la « logique inductive » proprement dite. Dans la version Carnap [1971], deux étapes supplémentaires s'insèrent entre la théorie normative et la logique inductive proprement dite.

La position méthodologique et philosophique la plus manifeste et la plus affirmée dans l'exposé de Carnap est le rappel constant de la distinction entre une théorie empirique et une théorie normative de la décision[21]. Il faut souligner l'importance de ce parti pris philosophique. Alors que les débats se prolongeront bien après Carnap entre deux positions qui sont parfois présentées comme des interprétations irréconciliables d'une même théorie, il est réconfortant de constater que Carnap affirme avec insistance qu'il s'agit de deux entreprises théoriques distinctes qui ont des objets distincts. Comme nous le soulignons à plusieurs endroits dans la présente thèse, il faut approuver la ténacité de Carnap à recommander l'utilisation de concepts différents. L'anti-psychologisme de Carnap, comme celui de son maître Gottlob Frege, est lié à une démarche qui vise à affranchir la théorie logique et analytique qu'il veut construire de toute forme de dépendance vis-à-vis des données empiriques et il est lié à la volonté d'éviter les difficultés posées par les concepts psychologiques. Il est clair que la théorie normative de la décision a un contenu psychologique moindre que la théorie descriptive. Alors que la théorie des décisions réelles (*actual decisions*) utilise le concept de croyance réelle (*actual belief*) que Carnap qualifie de psychologique, la théorie des décisions rationnelles utilise le concept de degré de croyance rationnelle. À la lumière des développements récents de la théorie du choix

rationnel, on est forcé de constater que ce second concept relève de l'épistémologie et que cette distinction ne semble correspondre à aucune différence dans le cas de croyances personnelles. À la suite de F. P. Ramsey, comme nous l'avons vu au chapitre II, on sait que l'on peut formaliser et définir quantitativement le degré de croyance réelle d'une personne par sa réaction d'acceptation ou de refus face à des paris qu'on lui propose. Or, Carnap adopte cette approche sans hésitations. Mais pour le degré de croyance rationnelle, Carnap propose la fonction de croyance d'un être parfaitement rationnel, un agent imaginaire et idéalisé. Il suggère ainsi une ramification des concepts dont l'effet secondaire serait d'éradiquer partiellement le psychologisme de la théorie du choix rationnel. Idéalement, dit-il en passant, il faudrait aussi distinguer entre « l'utilité réelle » et « l'utilité rationnelle »[22]. De Keynes à Savage, la tradition dont il s'inspire n'utilise pas de résultats psychologiques concernant le comportement des personnes pour construire les concepts de probabilité. Pour cette raison, Carnap estime à bon droit qu'il se situe dans le fil de la tradition. Dans la version de 1962, Carnap reproche à Bruno de Finetti d'utiliser l'expression « *subjective probability* » pour désigner les croyances réelles[23]. Enfin, un peu plus loin, il n'hésitera pas à désincarner entièrement son projet avec une référence à la robotique qui fait figure d'anticipation:

> Penser à la conception d'un robot nous aidera à trouver des règles de rationalité. Une fois découvertes, ces règles peuvent être appliquées non seulement à la conception de robots, mais aussi pour conseiller les êtres humains dans la prise de décisions aussi rationnelles que leurs capacités limitées le permettent.[24]

Mise à part la référence aux robots, cette citation est riche d'une autre observation pertinente. Dans la version de Carnap, la théorie normative de la décision idéalise l'agent tout en posant clairement qu'une des limites de la rationalité individuelle est déterminée par la nature de ses capacités cognitives limitées. On peut concevoir que ces limitations varient d'un agent à l'autre et renvoyer ce problème difficultueux au projet d'une théorie descriptive de la décision.

3.1 La définition fondamentale

Le principe qui gouverne la décision prend la forme d'un définition fondamentale qui combine l'utilité et la probabilité pour définir une relation d'ordre parmi les actes possibles. L'acte recommandé est celui qui maximise la valeur subjective *V*.[25] La formule qui indique comment calculer cette valeur est à tous égards habituelle. On remarque cependant que chez Carnap, la définition principale est formulée dans un langage entièrement explicite.

> Soit un agent *X* qui à un moment *t* doit faire un choix entre des actes possibles A_1, A_2,... . On suppose que *X sait* que les états possibles du monde (ou cette partie des états du monde qui est pertinente pour sa décision) au moment *T* sont W_1, W_2,... . Mais il ne sait pas quel état du monde est la réalité. Le nombre d'états du monde et le nombre d'actes possibles sont finis. S'il fait A_m et que l'état du monde est W_n, le résultat est $R_{m,n}$. Ce résultat $R_{m,n}$ est uniquement déterminé par A_m et W_n; de plus, *X* sait comment ce résultat est déterminé. On suppose qu'il existe

une fonction d'utilité U_X pour la personne X et que X connaît sa fonction d'utilité de façon à pouvoir calculer des valeurs en l'appliquant.

Sur cette base, on définit la *valeur subjective* (désirabilité) d'un acte possible, A_m pour l'agent X au moment T:

définition 3.1: $V_{X,T}(A_m) = \sum_n \left[U_X(R_{m,n})\, P(W_n)\right]$

où $P(W_n)$ la probabilité de l'état du monde W_n et où la somme porte sur tous les W_n.[26]

Pour la théorie des décisions réelles, la fonction P de la définition 1 s'interprète comme fonction de probabilité personnelle qui représente le *degré de croyance réel* de l'agent. Sur cette base, Carnap va construire les « étages supérieurs » en ajoutant des principes qui réduisent la classe des fonctions de croyance admissibles.

En progressant vers la logique inductive, P sera précisée, c'est-à-dire soumise à des contraintes additionnelles qui ont le statut d'exigences de rationalité. Sur le plan formel, Carnap conserve la forme de la définition 3.1. Cependant, il va proposer et discuter toute une famille de fonctions de probabilité pouvant occuper la place de P:

Cr : la fonction de croyance rationnelle;

Cr_n : la fonction de croyance relative aux états successifs de la sphère de croyance de l'agent;

Cred.: la fonction crédibilité (ou fonction de croyance initiale conditionnelle).

La forme générale de cette construction est originale. Ce n'est pas ainsi que les choses se présentaient dans les axiomatisations de la théorie de la décision vers 1950[27]. On peut élaborer une famille de théories en prenant la définition 1 comme base commune puisque c'est l'élément fondamental de la tradition néo-bernoullienne[28]. La construction étagée qui met en scène successivement les diverses fonctions de croyances (*P*, *Cr*, Cr_n, *Cred*) est particulière à Carnap; elle est une caractéristique significative sur le plan méthodologique. Parmi les mérites de l'analyse de la décision rationnelle par Carnap il faut souligner les clarifications méthodologiques et conceptuelles qui la structurent et qui restent pertinentes. En donnant aux postulats le statut de contraintes dont le rôle est de préciser et de donner de la substance au concept de rationalité, on pose l'explication du concept de rationalité comme un programme ouvert, qui ne se résoudra pas dans une fonction unique. Il y aura un grand nombre de façons de satisfaire ces contraintes, et donc un grand nombre de façons d'être rationnel, même pour des agents qui possèdent les mêmes informations — comme il y a plusieurs méthodes inductives. Mais, pour Carnap, ceci n'est vraiment admissible que pour la décision individuelle. Par contraste, la fonction de crédibilité est déjà plus impersonnelle et le programme de la logique inductive vise ultimement à construire un concept de rationalité qui ne soit pas personnel. En travaillant à formuler de telles contraintes, on donne de la substance au concept de croyance rationnelle et au concept de choix rationnel qui le présuppose. Sur le plan méthodologique, l'adéquation des contraintes de rationalité repose sur les tests que représentent la résolution de problèmes ou leur robustesse lorsqu'elles sont mises à l'épreuve par des expériences de pensées. À cet effet, le procédé utilisé pour mettre en difficulté les règles FR examinées plus haut est exemplaire.

> Les seules justifications que l'on peut fournir pour les exigences de rationalité sont nos jugements intuitifs concernant la validité inductive, c'est-à-dire concernant la rationalité inductive des décisions pratiques, à propos de certains paris. [29]

À nos yeux, cette validation ne présente aucune difficulté particulière et le développement de la logique de la décision montre que nous avons fréquemment des intuitions assez solides sur ce qui constitue un choix avantageux ou une offre raisonnable.

En conséquence de ce qui précède, on peut affirmer que la définition 1 fournit une plate-forme minimale sur laquelle on peut construire une logique de la décision dont les présupposés sont minimaux. Somme toute, c'est ce programme qui sera développé par R. Jeffrey, B. Skyrms et E. Eells.[30] Cette thèse doit néanmoins être clarifiée et justifiée. Même selon une interprétation purement normative, une théorie de la décision comporte des postulats qui limitent la classe des agents auxquels elle peut prétendre s'appliquer et la classe des mondes possibles dans lesquels ce type de rationalité peut exister. Ainsi, par exemple, dans le préambule de la définition 1, on suppose que pour le contexte où se pose le problème du choix, c'est-à-dire dans le « mini-monde » de l'agent, l'agent sait quels sont les états du monde qui résulteraient de ses actes. Il faudrait développer une explication philosophique de l'action et de l'intentionnalité qui montre que cette hypothèse n'est pas incompatible avec les capacités cognitives limitées de l'agent. En particulier, cette explication doit montrer que l'action intentionnelle implique que l'agent sait ce qu'il tente d'accomplir lorsqu'il cherche à faire en sorte que se réalise l'état du monde correspondant à $R_{m,n}$. Il faudrait aussi montrer que l'action

intentionnelle rationnelle implique que l'agent sait (et peut savoir) comment le monde sera modifié par ses actions; c'est ce qu'indique le préambule de la définition. L'agent peut faire cette prédiction sans que soit déterminé uniquement l'état du monde résultant de son action. Ainsi, son action peut avoir des effets secondaires imprévus qui, à la limite, peuvent empêcher que le résultat espéré se produise. Ce sont là des distinctions bien connues de la théorie de l'action et de l'intentionnalité et elles figurent nécessairement dans l'arrière-plan philosophique d'une logique de la décision. Une théorie de la délibération et de l'action rationnelle doit comporter une élucidation de l'action intentionnelle pour un agent dont les capacités sont limitées.

Un autre présupposé de la définition 3.1 exige un commentaire. Il ne serait pas raisonnable de supposer que cette définition explique le concept de valeur subjective d'un acte pour un agent dans un processus de délibération sans lui reconnaître la possibilité d'évaluer des probabilités et de comparer l'utilité de résultats envisagés. Nous avons vu ce problème au chapitre précédent et mentionné plus haut la méthode de Ramsey que Carnap accepte. Dans le même esprit, on peut évaluer l'utilité de la façon suivante. Lorsque j'évalue l'attrait d'une loterie, j'évalue tout à la fois l'intérêt du gain et la probabilité de l'obtenir. Si le fait de posséder tous les billets d'un tirage m'assurait d'obtenir un bien qui est accordé au gagnant, ce que je serais prêt à échanger contre l'avantage de posséder tous les billets mesure l'utilité du bien en question. Reste la question des probabilités: est-il bien légitime de supposer que l'agent possède des valeurs numériques précises pour P? Certainement pas; mais ce genre de théorie n'en exige pas tant. Écoutons plutôt Brian Skyrms:

> Une fausse conception fort répandue de la probabilité subjective doit être corrigée ici. C'est l'idée que la théorie dit que chaque personne rationnelle doit avoir une *valeur numérique précise* correspondant à son degré de croyance pour chaque proposition. Cela n'est pas vrai en réalité. [...] Tout ce qui est requis est que cette personne possède des préférences cohérentes qu'elle peut étendre à un domaine arbitrairement grand d'options.[31]

En d'autres termes, tout ce qui est exigé est que l'échelle d'évaluation des probabilités puisse se représenter par des nombres qui obéissent aux lois du calcul des probabilités. On peut illustrer la situation en reprenant une analogie que von Neumann et Morgenstern appliquent à la mesure de l'utilité. Comparons notre perception du probable à notre perception de la température ambiante.[32] Nous n'avons pas de valeurs numériques précises à faire correspondre à nos perceptions de la température ambiante. Néanmoins, si on peut le faire de façon cohérente, la relation *Ch (t,t')* =df « il fait plus chaud à *t* qu'à *t'* » qui figure dans nos jugements doit pouvoir se représenter sur une échelle dont les nombres se comportent comme ceux qui sont exprimés par les gradations qui figurent sur un thermomètre ordinaire. Nous savons tous qu'il serait assez facile de concevoir une expérience qui nous fasse poser des jugements incohérents sur la température ambiante. Il suffirait de nous faire circuler d'un sauna à une pièce tempérée, puis de nous faire séjourner dans une chambre froide avant de retourner dans la pièce tempérée. Telle autre expérience du même genre pourrait nous conduire à poser des jugements incompatibles avec la transitivité de la relation *Ch (t,t')*.

Il faut remarquer qu'on ne dirait certainement pas de telles expériences qu'elles *remettent en cause* la transitivité de *Ch (t,t')*. De plus, notre incapacité à indiquer des valeurs numériques précises ne serait pas retenue comme un

argument contre le concept exact de température, qui appartient de droit à la physique. En termes précis, nos jugements phénoménaux ne remettent pas en cause l'autorité normative de la conception physicaliste — métrique — de la température. Notre échelle d'évaluation doit pouvoir être représentée par des nombres qui obéissent aux lois de la métrique appropriée. Ici, l'idée de normativité tient entièrement dans cette exigence; elle doit s'appliquer à nos jugements de probabilité comme elle s'applique à nos jugements sur la température ambiante. Autrement, nos jugements ne sont pas fondés ou ils sont incohérents.

3.2 Quelques contraintes de rationalité

Nous allons poursuivre en examinant les principales contraintes qui sont considérées par Carnap, soit la cohérence, la cohérence stricte, la règle bayésienne de conditionalisation ainsi que la contrainte de symétrie qui s'applique à la *fonction de croyance initiale* Cr_0.

SR_1. Pour être rationnelle, *Cr* doit être cohérente

Une fonction de croyance *Cr* est dite *cohérente* si et seulement si, il n'existe pas de système de paris acceptables selon *Cr* et qui conduisent à une perte dans tous les cas possibles comme nous l'avons vu au chapitre II. C'est un résultat fondamental en théorie de la décision qu'une fonction de croyance est cohérente si et seulement si elle satisfait les axiomes de base de la théorie des probabilités. Une fonction *Cr* est cohérente si et seulement si elle est une mesure de probabilité normalisée.[33] La nécessité de tout ceci a fait l'objet

d'une démonstration attribuée à Bruno de Finetti à qui l'on doit cet usage de l'adjectif « cohérent ».

Un des axiomes de la théorie des probabilités prescrit que les énoncés tautologiques reçoivent la valeur 1. Dans le contexte de la décision rationnelle, Carnap ne donne pas d'indications particulières sur la façon dont les tautologies doivent s'interpréter dans le contexte d'une fonction de probabilité *personnelle*. La question ne va pas cependant pas de soi.[34] Ici, on peut interpréter la tautologie par ce que Carnap appelle une ***L-vérité***, c'est-à-dire, un énoncé qui est vrai dans toutes les situations possibles.[35] Si P est une proposition qui a la valeur 1, en symbole, $|P|=1$ la négation de P aura la valeur $1-|P|$. Par conséquent, seules les propositions dont la négation est impossible, c.-à-d. $1-|P|=0$, peuvent recevoir la valeur de probabilité 1. Du point de vue de la philosophie de Carnap ou dans un cadre théorique qui admet les jugements analytiques, tout ceci paraît bien raisonnable. Pour ne pas confondre ce qui est différent, c'est-à-dire pour ne pas confondre l'analyticité et la certitude, on n'assignerait pas une valeur $Cr\,(H)=1$ à une hypothèse contingente. Ce qui est plausible ici ne peut cependant pas être généralisé pour d'autres fonctions de probabilité discutées par Carnap. Dans le système de logique inductive, les choses seraient différentes. Il y a un axiome, dit d'auto-confirmation, qui pose que $C\,(E \mid E) = 1$ où C est le degré de confirmation, c'est-à-dire une fonction « purement logique » qui est la transposée de *Cred* et E est une proposition déjà connue.[36]

Carnap déplore que SR_1 soit la seule exigence que l'on retrouve chez la plupart des auteurs.[37] Il juge que les axiomatisations existantes sont extrêmement faibles et sous-déterminent l'explicandum à cause de leur généralité. Pour faire mieux, il reprend de A. Shimony le concept de

cohérence stricte.[38] Appelons moléculaires les énoncés ne contenant pas de quantificateurs; on dira d'une fonction *Cr* qu'elle est *strictement cohérente* si elle est une mesure de probabilité normalisée et qu'il n'y a pas de système fini de paris en accord avec *Cr* tel que, quoi qu'il advienne, le parieur ne peut faire un gain net mais il peut subir une perte dans au moins une éventualité. Dit autrement, un agent dont la fonction de croyance n'est pas strictement cohérente sera disposé à prendre des risques qui ne peuvent pas être compensés par des perspectives de gains. Carnap pose la cohérence stricte comme une seconde exigence fondamentale de la rationalité.

SR_2 Pour être rationnelle, une fonction de croyance doit être strictement cohérente.

Le critère de cohérence stricte empêche non seulement l'attribution d'une valeur *Cr (H)*=1 à un énoncé contingent mais aussi l'attribution de valeurs *Cr (H)*=0 et *Cr (H')*=1 aux énoncés contingents moléculaires. C'est ce qu'ajoute la définition de régularité. Il y a un théorème dû à J. Kemeny et R. S. Lehman qui montre que la régularité est une condition logiquement équivalente au critère de cohérence stricte. La différence essentielle entre le critère de cohérence et le critère de cohérence stricte tient à ce qu'une fonction de croyance *Cr* ne peut associer la valeur 0 à une hypothèse moléculaire contingente.

Comme nous l'avons déjà signalé, le fait de réserver les valeurs 0 et 1 aux propositions logiquement impossibles et aux tautologies respectivement est une approche qui peut se débattre sur le plan philosophique. Pour reprendre un exemple dû à Wesley Salmon, pourquoi ne pourrais-je avoir une

fonction de croyance qui associe la valeur de probabilité 0 à la possibilité de trouver un fil de cuivre qui ne conduise pas l'électricité? Je suis *certain* de ne jamais pouvoir trouver un tel fil, mais il n'y a pas d'impossibilité *logique* à concevoir son existence. En réalité, on peut considérer qu'il s'agit là d'une propriété du modèle qui est compréhensible et justifiable. Car en adoptant le point de vue bayésien, on peut contourner la critique en offrant une option de rechange. « Au lieu de dire que votre fonction de croyance donne la valeur 0 pour la proposition indiquée, dites plutôt, ou ce que vous voulez vraiment dire est... que vous êtes prêts à parier une fortune arbitrairement grande (c'est-à-dire aussi grande que vous le jugerez suffisant) contre presque rien que vous ne trouverez jamais un fil de cuivre qui ne conduit pas l'électricité ». Au prix d'enrégimenter un peu votre langage, vos intuitions peuvent être réconciliées avec les propriétés formelles de la cohérence et de la cohérence stricte. Une règle de paraphrase similaire pourrait être proposée pour les certitudes non-tautologiques.[39] La cohérence stricte demande que nous gardions l'esprit ouvert, disposés à admettre la possibilité de tout ce qui n'est pas nécessairement faux.

Pour Carnap, un système de postulats pour les probabilités personnelles ne devrait pas se restreindre aux cas des propositions nécessaires et des propositions impossibles. Pour aller plus avant dans la construction du concept de rationalité, Carnap propose d'examiner la façon dont l'agent acquiert et modifie ses croyances. Il remarque de façon perspicace que pour juger de la rationalité de quelqu'un, il faut non seulement considérer ses croyances (sa fonction *Cr*) mais il faut aussi considérer la façon dont il adopte et révise ses croyances.

3.3 La règle de révision des croyances

La troisième exigence de rationalité proposée par Carnap est la règle de conditionalisation bayésienne. Contrairement au formalisme de Carnap [1950] où les propositions étaient des entités intensionnelles exprimées par des énoncés, le formalisme utilisé ici définit les propositions comme des ensembles de points dans un espace de probabilité. La proposition *H* qui correspond à l'énoncé *h* est l'ensemble de points représentant les cas possibles où l'énoncé *h* est vrai. C'est ce qui explique qu'on retrouve l'opération d'intersection ensembliste plutôt que la conjonction dans la règle de révision:

SR$_3$. (a) La transformation de Cr_n en Cr_{n+1} dépend uniquement de la proposition E.

(b) Cr_{n+1} est déterminée par *Cr* et *E* de la façon suivante: Pour tout *H*,

$$Cr_{n+1}(H) = \frac{Cr_n(E \cap H)}{Cr_n(E)}$$

Cette règle exprime l'aspect cinématique de la fonction de croyance *Cr*. On l'appelle indifféremment « la règle de Bayes » ou « la règle de conditionalisation bayésienne ». Elle donne une méthode et pose des exigences quant à la façon de réviser nos croyances, c'est-à-dire de faire « évoluer » notre fonction de croyance. Elle a été abondamment discutée dans la littérature et possède maintenant plusieurs rivales.[40] On peut l'interpréter à l'aide de la paraphrase informelle suivante proposée par Salmon.

> Supposons que vous considériez l'hypothèse *H*. Avant de recueillir le prochain élément factuel, énoncez vos probabilités antérieures et vos prévisions. En enregistrant le prochain

> élément d'information factuelle, calculez vos probabilités postérieures pour l'hypothèse *H* en utilisant ces probabilités antérieures et ces prévisions. Modifiez maintenant votre degré de conviction en *H* en passant des probabilités antérieures aux probabilités postérieures.[41]

Les problèmes engendrés par cette règle ou les solutions envisagées pour la défendre ne sont pas particuliers au système théorique de Carnap. Pour cette raison nous nous limitons à quelques remarques ici. Un de ses inconvénients majeurs est de restreindre la révision des croyances aux seules acquisitions de résultats d'observations. En particulier, une conséquence immédiate de la clause (a) est que SR_3 ne permet pas la délibération à l'aide du raisonnement déductif! Ceci est un sérieux défaut. D'autre part, on peut montrer qu'elle permet parfois des révisions de croyances qui sont trop minimalistes pour être raisonnables. En fait, tout ce qui est requis est une redistribution des valeurs de probabilité personnelle qui soit compatible avec *H*.

On peut aussi montrer que la règle de révision exige que l'agent possède des probabilités antérieures non-nulles. Le théorème de Bayes ne permet pas la révision d'une hypothèse dont la probabilité antérieure serait 0 ou 1. Il faut donc supposer que l'agent possède, *avant toute observation*, une distribution de valeurs de probabilité qui les situent entre ces deux valeurs extrêmes. Cette distribution serait connue entièrement de façon a priori. C'est cette caractéristique — le problème des probabilité antérieures (*priors*) — qui est souvent donné comme un argument contre la logique inductive de Carnap.[42] On doit insister sur ce point qui peut relever de l'évidence pour un lecteur initié. Le problème des probabilités antérieures n'apparaît qu'avec la règle de conditionalisation. De plus, il n'y a pas vraiment de solution à ce problème

dans le cadre de la cinématique de la croyance de Carnap. C'est ainsi qu'après Carnap, d'autres auteurs parmi lesquels R. C. Jeffrey et B. Skyrms tenteront de dissoudre le problème sans trop réviser le bayésianisme. Ceci peut se faire en acceptant que les probabilités antérieures soient simplement une distribution arbitraire qui serait affinée par application successive de la règle de révision au cours de l'acquisition de nouvelles croyances. Les probabilités erronées se trouvent être éliminées, « lavées », par l'acquisition de nouvelles informations.

La modélisation de la révision rationnelle des croyances est un problème très étudié et qui a donné lieu à plusieurs développements ces dernières années. En nous faisant passer du domaine de la rationalité statique à celui de la rationalité cinématique, la règle SR_3 nous engage sur un chemin difficile. Pour nous, il constitue la limite du concept de choix rationnel au sein du programme plus vaste et peut-être trop ambitieux de la logique inductive. Dans la version Carnap [1971] la logique inductive est la sixième section d'un chapitre qui en compte sept. Seules les trois premières sections que nous venons de passer en revue portent sur *la décision* proprement dite. Procédant du schéma général de la prise de décision qui formait la première section, nous avons vu que Carnap discutait les *décisions réelles* dont l'étude forme ce qu'il nomme maintenant la théorie descriptive. Par la suite, il passait aux *décisions rationnelles* proprement dites pour lesquelles étaient énoncées les trois règles que nous venons de voir. Entre l'étude des décisions rationnelles et la logique inductive proprement dite, Carnap introduit les concepts de *crédibilité* et *de dispositions permanentes* qui nous font passer de la rationalité délibérative ou *prohairesis* à la rationalité cognitive.[43] Toujours en procédant par abstraction

et par généralisation, Carnap parvient à définir la base de connaissance empirique K_n, la croyance initiale conditionnelle $Cr_n(H)$ et la fonction de crédibilité, $CRED_X\ (H \mid K_{X,T})$, qui est en quelque sorte une version impersonnelle de *Cr*. Ces concepts débordent du cadre délimité par l'objectif que nous nous sommes fixé et qui était de circonscrire l'explication du choix rationnel chez Carnap. Ils opèrent la transition entre la problématique du choix rationnel qui relève d'une problématique de la rationalité pratique et la logique inductive proprement dite qui relève de la problématique de la rationalité épistémique.

3.4 Le principe d'indifférence et les axiomes d'invariance

De façon générale, lorsqu'on cherche à construire une explication d'une propriété, il est de grande utilité sur le plan méthodologique de préciser ce qui permettrait de juger que deux ou plusieurs individus possèdent également cette propriété. Dans le cas des probabilités, il est clair que ce concept est celui d'équipossibilité. Une définition de l'équipossibilité pourrait servir de point d'appui pour construire une explication du concept de probabilité. Or, il y a un principe bien connu qui semble fournir un critère opératoire de l'équipossibilité. Ce principe qui appartient à l'histoire de la théorie des probabilités était connu sous le nom de « principe de raison insuffisante ». Depuis Keynes [1921] qui l'a discuté et critiqué, on le connaît sous le nom de « principe d'indifférence ». C'est sous cette appellation qu'il est discuté par Savage et Carnap. Il y a plusieurs formulations de ce principe. Essayons d'abord de faire justice à la simplicité de l'idée qu'il exprime:

> *Principe d'indifférence*: Dans une situation d'ignorance complète concernant la vraisemblance de diverses possibilités, il est raisonnable de les considérer comme également probables.

Ce qui est remarquable dans ce principe est que son attrait intuitif se dissipe rapidement lorsqu'on réfléchit aux inférences qu'il rend légitimes. Si tout ce que l'on sait au sujet d'un dé à jouer est qu'il possède six faces, on ne peut pas en tirer la prédiction que les six faces apparaîtront avec une fréquence égale. Il se peut que le dé soit irrégulier ou plombé à notre insu. Keynes avait déjà critiqué et rejeté le principe d'indifférence alors qu'il tentait de construire une axiomatisation du calcul des probabilités. Carnap sait bien que ce principe ne vaut pas et sa position sur la question ne changera pas de Carnap [1950-4] à Carnap [1971].[44] Il croit cependant que le principe d'indifférence contient un noyau valide qu'il faut dégager et retenir. Ceci exige d'abord de bien saisir ce qui le rend invalide.[45] Il est assez clair, par exemple, que le principe d'indifférence n'est pas acceptable si on interprète le concept de probabilité en termes de fréquences. Mon ignorance ne pouvant pas affecter le résultat de jeter plusieurs fois un dé, la prédiction d'une fréquence égale pour les six faces du dé est absurde. Cependant, si on interprète les probabilités subjectivement, le principe semble retrouver sa valeur. Le principe nous autorise alors simplement à conclure que dans une situation d'ignorance complète sur la régularité d'un dé, il serait arbitraire d'avoir plus confiance que telle face apparaîtra plutôt que telle autre. Ainsi, malgré l'argumentation de Keynes, Jeffreys [1939] va accepter une variante du principe d'indifférence: « s'il n'y a aucune raison de croire à une hypothèse plutôt qu'à une autre, les probabilités sont égales ». Le problème est que, même dans

l'interprétation subjective des probabilités, le principe d'indifférence ne vaut pas. Pour s'en convaincre, il suffit de considérer le raisonnement suivant que propose Carnap.

Soit une urne dont on sait qu'elle ne contient que des boules bleues, jaunes ou rouges. La couleur des boules individuelles nous est inconnue de même que la proportion des boules de chaque couleur. Soit **B**, l'hypothèse que la première boule tirée de l'urne sera bleue, **R** l'hypothèse qu'elle sera rouge et **J**, l'hypothèse qu'elle sera jaune. Considérons les hypothèses **B** et non-**B**. Selon le principe d'indifférence, ces deux hypothèses ont la même probabilité, soit 1/2. Mais comme non-**B** est une proposition équivalente à « **R** ou **J** » qui sont équiprobables en vertu du principe d'indifférence, il s'ensuit que **R** a une probabilité de 1/4, de même que **J**. Les probabilités sont maintenant entièrement distribuées et **B**= 1/2, **R**=1/4 et **J**=1/4. Mais il y a un hic! Notre choix de considérer d'abord l'hypothèse **B** était arbitraire; si on avait d'abord considéré l'hypothèse **R**, l'hypothèse **B** aurait reçu une probabilité de 1/4 par le même raisonnement[46]. Donc, le raisonnement est incohérent et un système de postulats qui contient le principe d'indifférence, comme celui de H. Jeffreys est inconsistant.

Carnap considérait qu'une des questions fondamentales qui devait être résolue dans la construction d'un système de probabilités personnelles (inductives) était de trouver une version restreinte du principe d'indifférence qui soit acceptable. Dans *Logical Foundations of Probability*, il suggère de renoncer à définir le concept d'équipossibilité et propose de construire un système d'axiomes dans lequel ce concept est un terme primitif qui ne reçoit pas d'interprétation.[47] Dans les versions de Carnap [1963]b et de Carnap [1971], la logique inductive comporte des axiomes d'invariance, qui

constituent la partie valide du principe d'indifférence.[48] Ces principes énoncent qu'une méthode inductive (une fonction *c*) doit associer la même valeur *c(h,e)* pour toute permutation d'individus dans *h*. De même, la valeur de *c(h,e)* ne doit pas être affectée par une permutation des prédicats ou par une extension du domaine d'individus du langage pourvu que *e* ou *h* ne comportent pas de quantificateurs.

4. La logique inductive

Comme nous l'avons signalé au départ, l'explication du choix rationnel est pour Carnap une partie d'un programme plus vaste de construction d'une logique inductive qui explique la rationalité sur le plan épistémique et de façon quantitative. La logique inductive veut énoncer des lois qui gouvernent la formation et la révision des croyances à l'aide d'observations empiriques. Comme nous l'avons dit, elle appartient davantage à l'épistémologie qu'à la logique. On sait que Carnap a consacré les vingt-cinq dernières années de sa vie à ce programme et on retrouve les versions successives de sa logique inductive dans les textes et les ouvrages que nous avons discutés. Comme nous l'avons signalé, nous ne nous proposons pas de suivre Carnap sur ce terrain qui semble miné.

D'abord, même dans les petits modèles considérés par Carnap, il existe une infinité de fonctions *M* et de fonctions *C* qui correspondent à des méthodes inductives.[49] De plus, dans la logique inductive de Carnap, la décomposition du monde en possibilités élémentaires est relative à un langage. Or, un langage n'est pas neutre dans la façon dont il découpe le monde. Il y a fatalement une certaine quantité d'arbitraire qui s'insère dans ce découpage et

cet arbitraire est problématique pour une théorie qui se veut purement logique. Il y a aussi le problème des probabilités a priori que nous avons mentionné plus haut. Contrairement aux probabilités initiales d'une théorie bayésienne comme celle de Jeffrey qui s'effacent avec l'apprentissage et la révision de croyances, le probabilités *a priori* antérieures ne sont pas révisables. Enfin, la justification des axiomes de ces systèmes de logique inductive pose problème. Ces objections n'ont pas toutes la même valeur et l'analyse bayésienne du raisonnement scientifique est encore bien vivante comme le montre l'ouvrage de Howson et Urbach [1989]. Mais, il n'est pas exagéré de dire qu'il y a un consensus réel dans la littérature sur la non-viabilité de la logique inductive telle que la concevait Carnap. Dans la mesure où, exception faite de Salmon [1988], on identifie toujours l'explication de la rationalité chez Carnap et ses systèmes de logique inductive, il nous semble qu'on a négligé les aspects du probabilisme de Carnap que nous avons cherché à mettre en évidence.

Carnap critique également une conception de la logique inductive qui voudrait que celle-ci nous donne des *règles d'acceptation*. De telles règles nous indiqueraient s'il faut accepter telle ou telle hypothèse, la rejeter ou rester neutre face à elle. Or selon Carnap, de telles règles sont fatalement inadéquates.[50] Chez Carnap, le problème de l'induction devient le problème du degré de confirmation qui consiste à déterminer comment assigner des coefficients appelés degrés de confirmation à des énoncés qui ont le statut d'hypothèses. La logique inductive de Carnap s'inscrit dans un programme épistémologique empiriste qui est plus apparenté à la méthodologie hypothético-déductive de C. G. Hempel qu'au problème traditionnel de

l'induction. Le programme traditionnel de logique inductive, le « vieux problème de l'induction » est bien mort.

> Peu, s'il s'en trouve parmi les philosophes modernes, attendent encore la découverte de règles infaillibles pour faire des inférences inductives. En effet, avec de telles règles, nous pourrions acquérir une connaissance infaillible du futur, ce qui serait contraire à toutes nos croyances empiristes.[51]

Le programme proposé par Carnap pour la logique inductive rencontrera également sa part de difficultés. Mais l'évolution ultérieure de la théorie du choix rationnel voit renaître la problématique de l'induction sous d'autres formes. On peut la reconnaître, par exemple, dans le problème général de la révision des croyances: Quand devrais-je réviser mes connaissances ou ma fonction de croyance, et lorsque c'est pertinent de procéder à une révision, selon quelle règle parmi celles qui sont proposées? On peut considérer cette entreprise comme une héritière de la logique inductive. Elle se situe dans la foulée du problème de la révision discuté plus haut à la section 3.3. En généralisant une suggestion de I. Levi [1974], on pourrait soutenir que la logique inductive est plus vivante qu'il ne semble, et qu'on peut vraisemblablement extraire une *sorte* de logique inductive de tout modèle effectif de révision de notre sphère de croyance.[52] C'est le cas, par exemple, si une règle de révision valide une forme d'inférence ampliative.

Une des propriétés principales souhaitée par Carnap pour la logique inductive est qu'elle soit totalement indépendante de facteurs non-logiques. Ceci est une constante dans tous ses écrits sur la logique inductive, mais on trouve une insistance particulière dans Carnap [1963]b. Par contraste, les règles du choix rationnel font souvent référence à des éléments non-logiques

comme des gains ou des pertes.[53] Cette différence vient renforcer ce que nous affirmons, à savoir qu'il y a autonomie de l'explication du choix rationnel au sein de la logique inductive. Le caractère purement logique de la logique inductive est justifié explicitement à propos des fonctions *c(h,e)* qui représentent le degré de confirmation:

> Une fonction *c* est une fonction logique de ses arguments, c'est-à-dire que si une fonction *c* a été définie, alors sa valeur dépend uniquement des propriétés et des relations logiques ou sémantiques de ses deux arguments. De plus, les axiomes n'énoncent que des propriétés purement logiques des fonctions *c*. Donc, la théorie des fonctions *c*, basée sur les axiomes, est une partie de la logique.[54]

Il est bien clair que la logique inductive explique aussi un concept de rationalité: nous avons dit qu'il s'agit de rationalité épistémique plutôt que de rationalité délibérative (*prohairesis*). Nous ne nierons pas qu'une méthode inductive puisse être utile à un agent qui délibère. Ceci est une évidence: il en va de même pour tout autre ensemble de règles d'inférence qui permettraient d'enrichir de façon cohérente la sphère de croyance de l'agent et qu'il pourrait utiliser dans la délibération. Mais pour cette raison, la logique inductive n'a pas un statut privilégié dans la délibération.

5. La contribution de Carnap à l'analyse de la décision et de la délibération.

La décision rationnelle devient progressivement plus centrale dans la logique inductive de Carnap. On a vu que dans Carnap [1950-4], l'explication de la décision rationnelle était posée comme un exemple d'une application pratique

de la logique inductive. C'est la logique inductive qui expliquait le concept de probabilité$_1$. Mais déjà dans la seconde préface du même ouvrage qui date de 1962, Carnap propose une définition contextuelle du concept de probabilité qui s'appuie sur la règle de l'utilité espérée.

> Bien que le concept de probabilité au sens où nous l'entendons ici soit un concept purement logique, je crois que la signification d'énoncés comme « la probabilité de *h* relativement à *e* est 2/3 » peut être caractérisée en expliquant son usage, en combinaison avec le concept d'utilité, dans la règle de détermination des décisions rationnelles (vide supra, Règle FR$_5$). L'explication de la probabilité comme ratio d'un pari est un cas particulier simplifié de cette règle.[55]

Le contexte de cette citation indique clairement que par l'expression « ratio d'un pari », Carnap entend le ratio équitable d'un pari, c'est-à-dire une fraction dont le numérateur est le montant misé par le parieur et le dénominateur est la somme du montant misé et du gain espéré. Ainsi, pour une loterie dont le coût du billet est 2$ et le lot du gagnant est 1000 $, le ratio équitable du pari (*fair betting quotient*) est donné par la fraction 2$ ÷ (1000$ + 2$). La possibilité de définir ainsi les probabilités en termes de paris est un élément essentiel de la démonstration du théorème sur la cohérence de Bruno de Finetti, mieux connu sous le nom de « théorème du dutch Book ». Carnap avait déjà discuté différentes façons de définir le concept de probabilité$_1$ dans Carnap [1950], § 41. Les autres explications du concept de probabilité lui semblent cependant moins générales ou moins exactes.

Dans Carnap [1963]b, le choix rationnel occupe également une place centrale. En effet, une fois débarrassées de tout contenu psychologique, les

fonctions de probabilité sont identiques aux fonctions de crédibilité rationnelle. Ces dernières peuvent donc servir à valider la logique inductive.

> Un degré de confirmation est adéquat si, en l'utilisant comme fonction de croyance, il conduirait à une décision rationnelle.[56]

Cette citation témoigne d'un changement de statut important pour l'explication du choix rationnel. En effet, pour Carnap en 1950, il était possible d'espérer que le programme de la logique inductive pourrait s'édifier sur le concept de probabilité logique et que relativement à cette construction, l'explication du choix rationnel ne serait qu'un exemple d'application utile. Mais dans « Replies and systematic expositions » publié dans l'ouvrage de Schilpp [1963], il y a un retournement complet qui pose la règle du choix rationnel comme le concept fondamental de probabilité sur lequel repose la logique inductive et qui en fournit le critère de validation.

Comme nous l'avons vu dans ce chapitre, Carnap a contribué à valider l'interprétation subjective des probabilités sans rejeter pour autant l'interprétation fréquentiste. Il a soutenu l'importance de considérer la logique de la décision comme une explication du choix rationnel, ce qui est important sur le plan philosophique. La logique des probabilités personnelles n'est pas une simple règle de calcul pour manipuler les probabilités et une technique de résolution de problèmes. Enfin, Carnap a montré que le choix rationnel peut s'inscrire dans le projet plus vaste d'expliquer et de baliser les normes d'une épistémologie personnelle. R. Jeffrey a hérité de ce projet et son œuvre est un aller-retour incessant entre la logique de la confirmation et la logique de la décision. Du point de vue de la délibération et de la révision des croyances, la

différence principale est que dans la logique de la décision, c'est la délibération qui fait varier les valeurs de probabilité plutôt que l'observation.

CHAPITRE IV

LES MODELES STANDARDS DE LEONARD SAVAGE ET RICHARD JEFFREY

4.1. Introduction

Le but de ce chapitre est de formuler deux modèles standard de la logique de la décision en caractérisant certains aspects des théories de L. Savage et de R. Jeffrey. Nous utilisons l'expression « modèle standard » pour désigner ces versions modernes de la logique de la décision qui

(1) ont comme premier principe du choix rationnel une règle qui demande de maximiser l'utilité espérée (Savage) ou l'utilité espérée conditionnelle (Jeffrey) ;

(2) adoptent une conception subjective des probabilités ;

(3) s'interprètent principalement de façon normative.

En d'autres termes, on utilise l'expression « modèle standard » comme une abréviation de la longue expression qu'utilise Fishburn « modèle standard de l'utilité espérée subjective pour la prise de décision dans l'incertitude »[1]. L'expression « modèle standard » suggère aussi l'idée qu'il s'agit d'un point de référence, d'une base de comparaison et c'est bien de cela dont il s'agit dans le cas des théories de Leonard Savage et Richard Jeffrey. Dans ce chapitre, nous ne ferons pas un examen complet de ces théories et de leurs

conséquences connues ; nous allons en examiner quelques aspects de façon détaillée. Savage et Jeffrey ont donné des présentations relativement complètes de leurs théories dans des ouvrages qui sont accessibles pour des lecteurs non-spécialistes. En particulier, nous allons nous référer constamment à *The Foundations of Statistics* — ci-après, Savage [1954][2] — et à *The Logic of Decision* — ci-après, Jeffrey [1965]. Poursuivant notre objectif de préparer le terrain pour la seconde partie de la présente thèse, nous allons discuter les principaux concepts fondamentaux de ces théories ainsi que le modèle de la délibération qu'elles proposent car ce sont les aspects les plus pertinents pour la présente recherche. Savage et Jeffrey, en plus de faire connaître et de faire accepter l'interprétation subjective des probabilités, ont aussi fait connaître et accepter une épistémologie probabiliste qu'on appelle le bayésianisme. Il y a plusieurs variantes de bayésianisme, mais les thèses de base sur lesquelles s'entendent les bayésiens sont :

(i) qu'il faut tenir compte de ce que l'on sait ;
(ii) qu'il est naturel et utile d'exprimer nos connaissances dans le langage des probabilités ;
(iii) que si nos probabilités sont erronées, leur impact va s'effacer avec le temps.

La clause (i) est à l'opposé de l'épistémologie fondationnelle qui cherche des bases ultimes à la connaissance ou qui prend comme point de départ une table rase de toute connaissance. La clause (ii) exige la reconnaissance du langage des probabilités comme le seul langage légitime pour formuler les connaissances humaines. C'est ce qu'on désigne parfois sous le nom de

« probabilisme ». La clause (iii) indique que la classe de référence initiale relativement à laquelle on opère la révision de nos croyances peut être « contaminée » par des faussetés, car l'acquisition de nouvelles informations et la révision consécutive de notre sphère de croyance nous fera progresser vers l'élimination des faussetés.

Cette caractérisation du bayésianisme n'est pas très précise, mais elle suffit pour caractériser une orientation philosophique importante à laquelle Jeffrey et Savage sont fortement associés. Notons au passage que le type de bayésianisme auquel souscrit Jeffrey ne requiert pas que la fonction de probabilité qui assigne les probabilités de départ soit la fonction m^* de Carnap. Il peut être raisonnable de choisir une fonction de probabilité initiale qui soit impartiale, mais ceci n'est pas une obligation pour un agent bayésien qui délibère[3].

4.2 La théorie de Leonard Savage

Savage a élaboré sa logique de la décision à partir des conceptions de Ramsey, de Finetti et von Neumann et Morgenstern[4]. Il est clair que la théorie de Savage est très semblable à celle de Ramsey que nous avons examiné en détail au chapitre II. C'est certainement ce que pensait Savage car il écrit au sujet de *Truth and Probability* :

> Les concepts de probabilité et d'utilité de Ramsey sont essentiellement les mêmes que ceux qui sont présentés dans ce livre (Savage [1954]), mais le développement logique qu'il donne de ces concepts constitue une alternative intéressante au développement qui en est donné ici, ses définitions de la probabilité et de l'utilité étant simultanées et interdépendantes.[5]

Selon Luce et Raiffa [1957] la contribution de Savage à la théorie de la décision est de première importance et elle tient surtout au fait d'avoir unifié la théorie de l'utilité de von Neumann et Morgenstern et le calcul des probabilités subjectives élaboré par de Finetti[6]. Ainsi, nous retrouvons chez Savage, une logique de la décision dont plusieurs éléments nous sont déjà familiers car nous les aurons rencontrés aux chapitres précédents. Savage se fait le défenseur de la conception subjective (personnaliste) des probabilités, conception qu'il a largement contribué à faire connaître auprès des statisticiens en montrant la diversité de ses applications. Comme Ramsey, il annonce d'entrée de jeu que la logique de la décision qu'il expose est une « théorie abstraite du comportement idéalisé d'une personne rationnelle face à l'incertitude»[7]. Aussi, il apprécie de façon réaliste la valeur descriptive de sa théorie :

> Les décisions prises face à l'incertitude se retrouvent dans la vie de chaque individu et de chaque organisation. On pourrait même affirmer que les animaux prennent continuellement de telles décisions et que les mécanismes psychologiques à l'aide desquels les hommes prennent des décisions ont peut-être beaucoup en commun avec ceux qu'utilisent les animaux. Mais on peut présumer que le raisonnement formel ne joue aucun rôle dans la prise de décision chez l'animal, qu'il joue peu de rôle chez l'enfant, et moins qu'il ne serait souhaitable dans les décisions des hommes. On peut dire que l'objectif de ce livre, et en général, le but des statistiques, est de discuter des implications du raisonnement dans la prise des décisions[8].

Pour Savage, comme pour Ramsey et pour Carnap, l'incidence de facteurs psychologiques qui doivent demeurer en dehors d'une théorie formelle de la décision marque les limites de l'applicabilité du modèle à des situations

concrètes. Dans le même ordre d'idées, Savage défend la valeur normative de la logique de la décision de façon modérée :

> Lorsque certaines maximes vous seront proposées et que vous devrez en évaluer les mérites, vous devrez vous demander si vous-mêmes essayez de vous comporter selon ces maximes, ou, dit autrement, comment vous réagiriez si vous remarquiez que vous ne les respectez pas.[9]

On remarque que la première partie de cette citation suggère une interprétation descriptive des maximes de la théorie de la décision car il s'agit de savoir si nous utilisons certaines règles dans nos choix tandis que la seconde relève d'une interprétation plutôt normative. En m'interrogeant sur ma réaction alors que je remarquerais que je ne respecte pas une maxime de la théorie du choix rationnel, je m'interroge en fin de compte sur la valeur ou l'importance que j'accorde à une norme. Cependant, Savage ne rejette pas complètement l'interprétation descriptive de la théorie de l'utilité subjective comme en témoigne son essai sur l'observation des probabilités personnelles et les attentes, Savage [1971]. Dans cet article, il soutient que les *vraies personnes* sont comme des agents économiques idéaux possédant des préférences toutes faites face aux paris.

4.2.1 L'analyse d'un problème de décision

Dans les cinq premiers chapitres de *Foundations of Statistics*, on trouve une théorie axiomatisée du choix rationnel qui se distingue des autres théories par

quelques caractéristiques que nous allons examiner. La première de ces caractéristiques est la façon dont elle modélise et décompose un problème de décision. Pour Savage, un problème de décision se compose de trois éléments irréductibles qui sont encodés par trois concepts de base : les *actes*, les *états* du monde et les *conséquences*, ou résultats possibles[10] des actes d'un agent. Ce que Savage appelle le *monde* comprend tout ce qui peut nous intéresser dans une situation donnée. Dans un premier temps, un *état du monde* est défini comme une description complète du monde, ne laissant aucune indétermination sur un aspect pertinent[11]. On constate que Savage est plutôt explicite quant à l'ontologie formelle de sa théorie. Parmi les états du monde, Savage distingue *l'état véritable du monde* (*the true state of the world*) et réserve le terme *événement* pour désigner une collection d'états, un élément de 2^S. Ainsi, l'événement qu'il neigeait à Montréal hier à midi est un événement qui contient une infinité d'états. *L'événement universel*, noté S^*, est l'ensemble de tous les états du monde et *l'événement vide*, noté 0 est l'ensemble qui ne contient aucun état du monde. Dans la perspective de Savage, on a caractérisé un problème de décision lorsqu'on a précisé les *actions* parmi lesquelles l'agent pourrait choisir, les *résultats* (ou conséquences) de ces actions possibles et les *états du monde* qui déterminent le résultat que chaque acte produirait s'il était accompli. Nous utiliserons la notation qui suit pour les éléments d'un problème de décision :[12]

A est l'ensemble des actes $\{a, a', \ldots b, b', \ldots\}$;

C contient les résultats possibles $\{c, c', \ldots\}$ des actes ;

S l'ensemble des états du monde $\{s, s', \ldots\}$ qui résultent de chaque action dans ***A*** .

À ces termes, il faut ajouter la notation pour les événements, $\{E, E', \ldots\}$, qui sont des sous-ensembles de $\boldsymbol{S}$. Savage donne une caractérisation supplémentaire en disant que les actions sont des fonctions de S (les états du monde) dans C (les résultats). On peut donc utiliser la notation pour les actes $a(s)$ pour exprimer les éléments de $\boldsymbol{C}$. Cette conception des actes, qui les analyse comme des fonctions opérant sur des états, est un trait distinctif de la théorie de Savage. J. Joyce y voit une sorte d'axiome caché qui s'ajoute aux sept postulats de Savage[13]. C'est aussi une des hypothèses de base de sa théorie et on peut en questionner la signification. On peut se demander, par exemple, si les actes de Savage sont des actes faisables parmi lesquels l'agent peut choisir ou si ce sont plutôt des actes simplement imaginables que l'agent pourrait choisir, par exemple, dans des situations contrefactuelles[14]. En apparence, l'intérêt de cette définition fonctionnelle des actes est de nous fournir un critère d'identité des actes qui semble absolument clair. Deux actes sont identiques s'ils produisent les mêmes résultats pour les mêmes états. Cependant, il y a plusieurs façons de décrire le domaine de cette fonction. Ainsi, l'explication de Savage n'écarte pas toute ambiguïté et cette question donne lieu à des débats d'interprétation. Savage indique aussi que ce qu'il appelle un état du monde est une description complète du monde contenant chaque aspect pertinent[15]. Mais la nature exacte de ces éléments premiers n'est pas expliquée très précisément par Savage ; par conséquent, on doit les considérer comme termes primitifs dans sa théorie. Le rôle de chacun des facteurs d'un problème de décision est facile à cerner. Les états du monde sont l'objet des croyances et des incertitudes de l'agent ; les états du monde reçoivent les valeurs de probabilité. Les résultats ou conséquences servent à exprimer ce que l'agent désire de manière non instrumentale, autrement dit,

des biens désirables pour eux-mêmes. Les résultats sont comparables entre eux par la relation de préférence. Enfin, les actions sont envisagées comme étant désirables de façon instrumentale et non pour elles-mêmes[16].

La manière d'analyser un problème de décision est un des traits distinctifs de la théorie de Savage. Faisons une courte digression pour contraster cette base conceptuelle avec celle de la théorie de Ramsey que nous avons discutée au chapitre II et celle de Jeffrey que nous étudierons plus loin dans ce chapitre. Chez Ramsey, les degrés de croyance, autrement dit, les valeurs de probabilité sont attribuées aux propositions tandis que chez Savage, elles sont attribuées aux événements. Comme nous l'avons signalé au chapitre II, Ramsey admet avoir en tête une théorie des propositions du même type que celle de Wittgenstein. La théorie de Ramsey admettait comme terme primitif la notion de *monde*, et les mondes jouaient chez Ramsey simultanément les rôles de *résultat d'une action* et de *valeur du résultat d'une action.* L'action n'était pas représentée an tant que telle dans la théorie de Ramsey. Dans la théorie de Ramsey, il serait vrai de dire qu'agir, c'est actualiser un monde ou, comme nous le disions, une famille de mondes. Il faut nuancer un peu cette affirmation pour tenir compte de la dimension probabiliste et ajouter que les options qui s'offrent à un agent dans un problème de décision sont en réalité des loteries sur des mondes. Une autre différence notable entre la théorie de Ramsey et celle de Savage est que la théorie de Savage se passe entièrement du concept de proposition éthiquement neutre. On peut dire que la logique des préférences chez Savage est entièrement structurale et qu'elle évite le problème de calibration qui préoccupait Ramsey. En anticipant un peu sur ce que nous dirons plus loin, on peut aussi contraster la logique de la décision de Savage avec celle de Jeffrey. Pour ce dernier, il est possible d'unifier les

concepts de base de la théorie de la décision en utilisant comme seul type primitif le concept de proposition. Actions, états du monde, résultats des actions, ces éléments peuvent tous être exprimés et définis en termes de propositions. Il serait abusif d'accuser Jeffrey de chercher à nier la différence entre les divers éléments distingués par la théorie de Savage. Il s'appuie et utilise le fait qu'une même proposition telle que « ma voiture est en panne » devient selon les contextes le déterminant d'un état du monde, le résultat d'une action ou l'objet d'une croyance. Cet objectif d'unification, Jeffrey l'énonce très clairement dans Jeffrey [1992]b. Il s'agit d'unifier le type d'entité auquel on attribue les valeurs de probabilité et les valeurs d'utilité. Il y a bien une différence théorique notable entre la théorie de Jeffrey et celle de Savage. Pour Jeffrey, toute proposition peut être le contenu possible d'une croyance, d'une préférence (un désir) alors que pour Savage seuls les états peuvent être l'objet de croyances, les résultats et les actes pouvant faire l'objet de préférences. Il n'est pas si évident d'apprécier l'importance théorique de cette différence. Notons en passant que Jeffrey utilise le concept de proposition dans un sens assez ordinaire. Autrement dit, pour formuler une théorie de la décision comme la sienne, il faut utiliser un concept de proposition qui vérifie quelques propriétés habituelles. Rappelons quatre de ces propriétés. (1) Elles sont vraies ou fausses ; (2) on forme des conjonctions de propositions pour décrire des états de chose ; (3) les propositions peuvent être combinées par les connecteurs propositionnels et modifiées par des opérations comme la négation ou les modalités ; (4) les propositions peuvent servir à représenter ce qu'on appelle les objets (ou contenus) des croyances et des préférences des personnes. Plusieurs théories des propositions peuvent satisfaire les contraintes que nous venons d'énoncer. En regard de la clause

(4), on pourrait vouloir utiliser une conception fine des propositions, une conception capable d'offrir une solution efficace au problème de l'opacité des contextes engendrés par les formes « je crois que *p* » ou « je préfère *p* à *q* ». Cependant, on remarque que les difficultés associées au fait de prendre les propositions comme objets des croyances ne sont pas discutées en logique de la décision, comme si l'énigme posée par les attitudes propositionnelles n'avait aucun impact sur l'analyse et l'interprétation des problèmes de décision. Tout bien réfléchi, d'un point de vue strictement opératoire, cette position est tenable, même dans une perspective critique et réfléchie. Elle est tenable parce que tout dépend de l'interprétation et de la visée théorique que l'on donne de la logique de la décision que l'on cherche à construire. Plusieurs théoriciens, mais certainement pas Jeffrey, considèrent cette théorie comme une technique mathématique qui se valide de façon opératoire, c'est-à-dire une méthode de résolution de problèmes. Sans souscrire à cette interprétation, on peut cependant en dériver une solution efficace et raisonnable au problème de la sensibilité aux formulations. En effet, dans cette perspective, un problème bien posé en théorie de la décision devrait recevoir une formulation qui évite soigneusement les énigmes logiques ou sémantiques. Il arrive parfois qu'un problème soit difficile à analyser ou à résoudre tout simplement parce qu'il est mal posé. Dans cette optique, on peut considérer qu'il n'y a pas d'enjeu théorique significatif à adopter une théorie des propositions qui admet la clause (4) dans le contexte de ce type de théorie de la décision. Il faut aussi remarquer que Jeffrey ne voit pas d'objection à interpréter les propositions comme des énoncés à la manière de Quine. Ceci est conforme à sa perspective philosophique d'orientation empiriste.

De façon plus générale, on peut s'interroger sur l'importance de la différence dans le choix des termes primitifs. Certaines théories prennent comme terme primitif le terme « d'événement », d'autres le terme « proposition». Comme le note Isaac Levi, Savage aurait pu exprimer sa théorie en parlant de *descriptions d'actes*, de *descriptions d'états* et de *descriptions de conséquences*[17]. Dans la mesure où il est possible de reformuler les théories du premier type dans le langage des théories du second type et réciproquement, on peut soutenir que la différence n'est pas significative. Du point de vue de la méthodologie, on appelle théorie un ensemble d'énoncés fermé sous une relation de conséquence et on dit que deux théories qui sont traduisibles l'une dans l'autre sont logiquement équivalentes. Bien sûr, il s'agit d'un sens extensionnel faible de l'équivalence entre théories[18].

4.2.2 L'axiomatisation de Savage

Les postulats de Savage et son théorème de représentation forment une théorie qui a été reprise, développée ou critiquée par plusieurs auteurs. La formulation que nous donnons ici s'appuie sur diverses discussions critiques des postulats de Savage parmi lesquelles nous mentionnons Joyce [1999], Picavet [1996], Shafer [1986] Fishburn [1964], [1970] et [1981], ainsi que Luce et Raiffa [1957]. Il y a plusieurs différences entre ces formulations de la théorie de Savage et il n'est pas facile de retracer la correspondance des postulats. Bien qu'elle ne soit pas strictement équivalente à la théorie de Savage, on considère habituellement l'axiomatisation de Aumann et Anscombe [1963] comme une version simplifiée de l'axiomatisation de Savage. Comme

certaines des loteries de Aumann incluent des probabilités objectives, il nous a semblé qu'elle était plus une lointaine cousine qu'une sœur jumelle de la théorie de Savage[19]. Parmi les sources secondaires que nous avons utilisé, mentionnons que Joyce [1999] se distingue parce qu'on y trouve une formulation actualisée de la théorie de Savage qui la rapproche de celle de Jeffrey, que Luce et Raiffa [1957] ainsi que Picavet [1996] sont des exposés accessibles au non-initié tandis que Shafer [1986] est une critique qui propose une révision substantielle de la théorie de Savage. Notre formulation des postulats de Savage reprend la relation de préférence ($\preceq$) telle qu'elle se trouve dans Savage [1954], elle capitalise sur la simplicité de certaines formulations dues à Luce et Raiffa, et met à profit certaines caractéristiques de l'élégante formulation de Fishburn [1981]. À notre avis, ce sont les discussions approfondies de Fishburn et de Joyce qui fournissent les analyses les plus étoffées.

Le but de Savage était d'analyser explicitement et de justifier la théorie de l'utilité espérée en établissant un ensemble de postulats qui s'interprètent comme un ensemble de contraintes dont on peut montrer que seul un agent qui maximise l'utilité espérée va les respecter. Comme toute théorie de l'utilité espérée, celle de Savage comporte une analyse du concept de préférence. Ici, le point de départ est la relation qui correspond à la fonction propositionnelle « ...n'est pas préféré à... » et elle sera notée $\preceq$. Dans l'interprétation visée, cette relation est une forme particulière de la relation de préférence. Pour la distinguer d'autres façons d'exprimer la relation de préférence, nous utiliserons l'expression *relation de non-préférence* pour désigner $\preceq$. Cette relation définit un ordre simple (on dit aussi un préordre complet) sur les actes. Formellement, elle possède les propriétés suivantes pour tout x, y, z :

Connexité[20] : pour tout couple (x, y), on a : $x \preceq y$ ou $y \preceq x$

Transitivité : $(x \preceq y$ et $y \preceq z) \rightarrow x \preceq z$

Comme nous le verrons, la relation de non-préférence est d'abord définie sur les actes, puis sur les actes mixtes, les conséquences et enfin les paris.

À la suite de cette caractérisation, Savage peut définir quelques autres relations connexes :

la relation d'indifférence, notée …≈…: $a \approx a' =_{df} a \preceq a'$ et $a' \preceq a$

la réciproque de $\preceq$, notée $\succeq$: $a \succeq a' =_{df} a' \preceq a$

l'acte a' est préféré à a :$a \prec a' =_{df}$ il est faux que $a' \preceq a$

la préférence stricte, notée …≻… : $a \succ a' =_{df} a' \prec a$

Savage rejette explicitement la tentation (*sic)* d'analyser la préférence comme une relation d'ordre partiel[21]. Pour ce faire, on remplacerait la connexité par la réflexivité tout en laissant la possibilité que certaines paires soient incomparables. Il est douteux selon Savage que cet affaiblissement de la relation de préférence entraîne un gain théorique quelconque.

Les propriétés de la relation $\preceq$ sont fixées par le premier postulat de l'axiomatisation de Savage ainsi que par la définition qui l'accompagne :

P_s 1 : La relation $\preceq$ est une relation d'ordre simple sur les actes.

En commentant ce postulat, Savage profite de l'occasion pour signaler qu'on peut l'interpréter, comme chacun des autres postulats de sa théorie, de façon normative ou empirique (descriptive). Si la logique devait être interprétée de façon descriptive dit-il, il faudrait la comprendre comme une théorie

approximative et grossière alors que sa valeur principale se révèle dans l'interprétation normative.

D_s 1 : La signification de l'expression $a \preceq a'$ *étant donné* E est que si a et a' sont modifiés pour devenir respectivement b et b', de telle façon que les conséquences de a sont les mêmes que celles de b et celles de a' sont les mêmes que celles de b' pour tous les états s qui ne sont pas dans E, alors, pourvu que b et b' coïncident avec a et a' dans E — ce qui signifie que leurs conséquences sont les mêmes étant donné E — il s'avère que $b \preceq b'$.[22]

La définition D_s1 n'étant pas évidente par elle-même, ajoutons un bref commentaire. Comme le note Savage, l'idée principale de cette définition est que le sens de l'expression définie ici ne devrait pas dépendre des valeurs de a et de a' en dehors de l'événement E. Les actes b et b' sont des projections des actes a et a' ; ils sont modifiés pour coïncider dans les états qui sont à l'extérieur de E. Ainsi, pour que la définition de la préférence conditionnelle soit exacte, il faut que la relation de préférence sur les actes modifiés b et b' ne dépende pas de leurs conséquences pour les états qui sont en dehors de E[23]. C'est pourquoi Savage pose qu'ils coïncident.

L'axiome qui suit énonce un principe *d'indépendance*. Il s'agit bien d'indépendance, parce qu'étant donné un couple d'actes ordonné par la relation de préférence, on peut déduire un autre couple d'actes ordonné par la même relation aux conditions que nous venons d'énoncer dans la définition de la préférence conditionnelle. En quelque sorte, ce deuxième postulat affirme que la définition donnée pour la relation de préférence conditionnelle est

adéquate parce que d'un ordre de préférence, on peut en déduire un autre. Avant d'énoncer son second postulat, Savage énonce un principe qu'il qualifie « d'extra-logique », le principe de la *chose sûre* (*sure-thing principle*). Savage renonce à l'idée de formuler précisément le principe de la chose sûre car il faudrait utiliser dans la définition des termes comme « savoir » et « possibilité », et ces termes théoriques doivent en fin de compte demeurer indéfinis[24]. On ne peut s'empêcher de remarquer que ces propos de Savage remontent à près d'un demi-siècle et qu'ils sont moins justifiables aujourd'hui. En effet, la logique philosophique offre de nos jours des langages pragmatiques désambiguïsés qui contiennent des opérateurs modaux et épistémiques que l'on peut doter d'une sémantique rigoureusement explicite. Dans notre recherche, nous avons adopté comme visée méthodologique d'analyser la logique de la décision en tenant compte des avancées de la logique philosophique contemporaine. Une partie de la tâche qui nous incombe dans notre effort pour réaliser cette analyse est de rendre explicite les présupposés modaux et épistémiques qui sont habituellement laissés dans l'ombre et de tirer les conséquences théoriques de ce cet éclairage. Nous discuterons plus explicitement cette question au chapitre VI.

Commençons par introduire une notation pour l'opération de complémentation ensembliste ; comme Savage, nous utilisons l'expression « $\sim X$ » pour abréger « le complément de X ». Le principe de la *chose sûre* énonce que si une personne ne préfèrerait pas a à a', en sachant que la condition B est réalisée ou en sachant qu'elle n'est pas réalisée — ci-après $s \in \sim B$ — , alors il ne préfère pas a à a'. Ainsi, si je ne préfèrerais pas *la possibilité* d'aller faire une promenade plutôt que de rester à la maison, *sachant* qu'il pleut ou sachant qu'il ne pleut pas, alors je ne préfère pas faire

une promenade plutôt que rester à la maison. En quelque sorte, le principe de la chose sûre supprime la relativité de la préférence à la condition B.

P_s2 : Si *a*, *a'*, *b* et *b'* sont des actes qui satisfont les conditions suivantes :

1. Hors de l'événement *B*, *a* coïncide avec *a'* et *b* coïncide avec *b'*,
2. Dans *B*, *a* coïncide avec *b* et *a'* coïncide avec *b'*,
3. $a \preceq a'$;

alors $b \preceq b'$.

Le postulat P_s2 peut servir de modèle pour définir de manière analogue les relations d'indifférence conditionnelle, $a \approx a'$ *quand B* et de même pour les relations conditionnelles $\succeq$, $\prec$, et $\succ$. Notons au passage que ce postulat est la cible d'une critique très célèbre due à Maurice Allais et nous examinerons cette critique du principe d'indépendance dans la section suivante avec d'autres critiques des postulats de Savage[25].

Avant de formuler le troisième postulat, nous allons formuler deux définitions incontournables. Tout d'abord, nous appellerons *actes constants*, les actes qui ont les mêmes conséquences pour tous les états où ils sont accomplis. Dans ce qui précède, la relation « ...n'est pas préféré à... » est une relation d'abord définie sur les actes. La définition du concept d'acte constant permet d'exporter la relation $\succeq$ pour qu'elle s'applique aussi aux conséquences. Une *conséquence c* est définie comme le résultat de l'acte constant qui la produit pour chaque état *s* [26]. C'est ce qu'exprime la définition $D_s 2$.

D_s2 : Soit $a\,(s) = c$ et $a'(s) = c'$ pour chaque état s dans S, alors nous dirons que $c \succeq c'$ si et seulement $a \succeq a'$. Dans la clause qui précède, les actes a et a' sont dits *constants*.

Vient ensuite la définition de l'événement nul ;

D_s3 : On dit d'un événement ϕ qu'il est un *événement nul* si toutes les paires d'actes sont indifférents étant donné ϕ, autrement dit, $a \succeq a'$ et $a' \succeq a$ pour tout a, a' étant donné ϕ.

En réalité, comme le remarque Joyce, cette définition exige que l'agent soit cohérent dans son appréciation des événements nuls. Elle va permettre aussi, avec l'aide des autres postulats, de montrer que l'événement nul reçoit la probabilité 0 pour toute fonction de probabilité qui représente les croyances de l'agent[27].

P_s3 : Si E n'est pas un événement nul et que $a\,(s) = c$ et $a'(s) = c'$ pour chaque état s dans E, alors $a \succeq a'$ étant donné E si et seulement si $c \succeq c'$.

Le postulat P_s3 énonce que les préférences conditionnelles pour des actes n'affectent pas les préférences pour des conséquences. Autrement dit la relation des actes aux conséquences est monotone sur l'ensemble des événements. Ce postulat énonce aussi un principe « d'admissibilité » qui correspond au rejet des actes dominés.

D_s4 : On dira qu'un événement *E n'est pas plus probable* qu'un événement E' s'il suffit que les trois conditions suivantes soient satisfaites pour que $a \succeq a'$

1) soit c et c', une paire quelconque de conséquences telles que $c \succ c'$

2) $a\,(s) = c$ pour l'état s dans E et $a'(s) = c'$ pour s en dehors de E

3) $a'(s) = c$ pour s dans E' et $a'(s) = c'$ pour s en dehors de E.

Le postulat suivant est exprimé de façon fort différente dans les diverses reformulations de la théorie de Savage. Selon Luce et Raiffa [1957] le quatrième postulat de Savage ajoute simplement à la définition D_s4 l'idée que du point de vue des probabilités, toute paire d'événements est comparable, c.-à-d. pour toute paire d'événements $< E, E' >$, $P\,(E) \geq P\,(E')$ ou $P\,(E') \geq P\,(E)$.[28] En réalité, la fonction du quatrième postulat de Savage est d'indiquer comment découvrir les probabilités (les degrés de croyance) à partir des préférences. Sans chercher à tout prix la définition la plus courte, nous allons suivre et reformuler l'explication de Savage. Ce dernier introduit d'abord la définition d'un pari. Un *pari* au sens de Savage, c'est-à-dire, *offrir un prix* à quelqu'un si E est réalisé consiste à lui donner la possibilité de faire l'acte a étant donné E, ici abrégé par le terme a_E, tel que

$a_E\,(s) = a$ pour s dans E, et

$a_E\,(s) = a'$ pour s en dehors de E

où $a' \prec a$

Sur cette base, Savage introduit le postulat de dominance stochastique qui comporte quatre clauses comparables à la définition précédente[29] :

P_s4 : (dominance stochastique) Si a, a', b et b', des actes ; E, F des événements et les actes conditionnels a_E, a_F , b_E, b'_F , sont tels que

(1) $a' \prec a$ et $b \prec b'$

(2) $a_E(s) = a$, $b_E(s) = b$, pour $s \in E$

$a_E(s) = a'$, $b_E(s) = b'$, pour $s \in \sim E$

(3) $a_F(s) = a, \quad b_F(s) = b, \quad$ pour $s \in F$

$a_F(s) = a', \quad b_F(s) = b', \quad$ pour $s \in \sim F$

(4) $a_E \preceq a_F$

alors $b_E \preceq b'_F$

Comme les actes a_E, a_F, b_E, b'_F que l'on compare sont des paris, le postulat P_s4 énonce que la préférence pour un pari sur un autre n'est pas modifiée lorsque les prix associés à ces paris sont modifiés sans que la relation entre ces prix soit modifiée. La relation de préférence étant ainsi étendue aux paris sur E ou sur F, il devient possible d'ordonner les événements E et F par la relation de probabilité « $\ldots \leq \ldots$ ». On dira que l'événement E n'est pas plus probable que l'événement F, $E \leq F$ si et seulement si

$a_E(s) = a$ pour $s \in E$ et $a_E(s) = a'$ pour $s \in \sim E$

$a_F(s) = a$ pour $s \in F$ et $a_F(s) = a'$ pour $s \in \sim E$

La méthode utilisée pour définir le concept central de probabilité personnelle — méthode qui est commune à Ramsey, Savage et Jeffrey — revient ni plus ni moins qu'à dériver les valeurs de probabilité des préférences. Le postulat suivant exprime une condition de non-trivialité qui impose une contrainte minimale sur les préférences de l'agent.

P_s5 : Il existe au moins une paire d'actes qui ne sont pas indifférents.

Autrement dit, il y a au moins une paire a et a' tels que $a \prec a'$ ou $a' \prec a$.
Rappelons qu'une *partition* de S est un ensemble d'événements disjoints de S dont l'intersection est vide et dont l'union est égale à S. En d'autres termes, $E \cap E' \cap E' \ldots = \Lambda$ (Λ := l'ensemble vide) et $E \cup E' \cup E'' \ldots = S$.

P_s6 : Supposons que $a \succ a'$. Pour chaque conséquence c, peu importe la désirabilité de c, il y a une partition suffisamment fine de S en un nombre fini d'événements tels que si a ou a' sont modifiés pour produire le résultat c pour chaque événement de la partition, alors a demeure préférée à a'.

Le postulat P_s6 est une condition de continuité. Il est nécessaire pour démontrer le théorème d'existence d'une mesure de probabilité définie sur l'ensemble des événements. C'est le théorème T_s1 qui figure à la page suivante. Comme le note Fishburn, l'effet de ce postulat est d'interdire que la structure des préférences de l'agent comporte des conséquences infiniment désirables ou infiniment indésirables. Il indique clairement que l'ensemble S est de cardinalité non dénombrable. Ce trait a donné lieu à des critiques sur l'applicabilité de la théorie, c'est-à-dire son adéquation descriptive. Savage comprenait bien la difficulté et il observe que la restriction à un domaine fini compliquerait considérablement la théorie. En utilisant ce critère de simplicité pour justifier P_s6, Savage formule un plaidoyer efficace pour l'usage de l'infini en mathématique[30]. Notons au passage que l'inexistence des conséquences infiniment désirables entraîne qu'un agent qui satisfait les postulats de Savage ne peut se retrouver dans un paradoxe de Saint-Petersbourg. Autrement dit, on ne peut pas construire une *pompe à fric* (*dutch book*) pour un agent *à la Savage* en lui faisant miroiter un gain infini.

P_s7 : Soit l'acte a' ; appelons a'_s l'acte constant qui coïncide avec a' pour l'état s. On a

(1) si $a \succeq a'_s$ étant donné E pour tout s dans E, alors $a \succeq a'_s$ étant donné E ; et

(2) si $a'_s \succeq a$ étant donné E pour tout s dans E, alors $a'_s \succeq a$ étant donné E [31].

Fishburn observe que ce dernier postulat exprime une condition de dominance. Savage montre qu'il est indépendant des postulats P_s1 à P_s6 et utilise ce postulat pour étendre le concept d'utilité à des actes qui ont un nombre infini de conséquences.

À partir de ces sept postulats et d'eux seuls, Savage est en mesure de prouver les deux théorèmes, T_s1 et T_s2, qui garantissent respectivement l'existence d'une *mesure de probabilité personnaliste* (subjective) (T_s1) et l'existence d'une *fonction d'utilité* (T_s2). Comme chez Ramsey, cette fonction d'utilité est unique pour la classe d'équivalence engendrée par les transformations linéaires[32]. Nous allons énoncer ces théorèmes pour pouvoir nous y référer par la suite[33].

T_s1 : Il existe une fonction à valeurs réelles unique P, appelée *mesure de probabilité personnaliste*, définie pour l'ensemble des événements (sous-ensembles de E) et telle que

(1) Pour tout E, $P(E) \geq 0$;

(2) $P(S^*) = 1$;

(3) Soit E et E', deux ensembles disjoints, alors

$$P(E \cup E') = P(E) + P(E') ;$$

(4) E n'est pas plus probable que E' si et seulement si $P(E) \leq P(E')$.

T_s2 : Il existe une fonction à valeurs réelles u, appelée *fonction d'utilité*, définie sur l'ensemble des conséquences ayant la propriété suivante : Si E_i, pour $i = 1,2,\ldots, n$, est une partition de S et a est un acte ayant la

conséquence c_i sur E_i et si E_i' pour $i = 1,2,\ldots, m$, est une autre partition de S et que a' est un acte ayant pour conséquence c_i' sur E_i', alors $a \succeq a'$ si et seulement si

$$\sum_{i=1}^{n} u(c_i)\, P(E_i) \quad \geq \quad \sum_{i=1}^{m} u(c_i')\, P(E_i')$$

Ceci complète notre exposé de la théorie de Savage. Dans la prochaine section, nous allons formuler quelques commentaires généraux et examiner quelques-unes des principales critiques adressées à cette théorie.

4.2.3 Commentaires et perspectives critiques

Comme premier commentaire, nous allons appliquer au système de postulats de Savage la distinction entre les axiomes de structure et les axiomes de rationalité. On se souviendra qu'en exposant la théorie de Ramsey au chapitre II, nous avons remarqué qu'il était possible de distinguer ces deux types d'axiomes. Rappelons que les axiomes de structure expriment des propriétés formelles que l'on peut interpréter comme des exigences d'adéquation formelle de la théorie tandis que les axiomes de rationalité expriment des exigences qui définissent le choix rationnel pour un agent. Nous avions aussi remarqué qu'il était plus facile de tracer cette distinction abstraitement que de l'appliquer. Comme les axiomes de structure sont des artifices du modèle plutôt que des contraintes de rationalité, on sait *a priori* qu'il n'y a pas de débats substantiels qui les concernent. Selon nous, on évite ainsi de faux

débats car en séparant les deux types de postulats, on délimite le domaine des critiques potentiellement pertinentes.

Parce qu'ils imposent clairement des contraintes sur la structure de préférence d'un agent P_s1, P_s2, P_s3 et P_s4 sont des axiomes de rationalité. Cependant, il faut noter que Joyce scinde le postulat P_s1 en deux postulats distincts pour distinguer d'une part, les clauses qui caractérisent l'ordre des préférences, qui sont des axiomes de rationalité et d'autre part, la propriété de connexité elle-même dont la raison d'être est structurale. Pour fixer le statut du postulat P_s4, il faut appliquer le critère avec soin. P_s4 : est-il un énoncé existentiel qui concerne la taille et la complexité de la structure des préférences de l'agent ? Est-il plutôt un énoncé universel qui n'impose aucune contrainte sur la structure des préférences de l'agent, mais qui doit être satisfait pour garantir la représentation en termes de probabilité ? Pour comprendre le statut de P_s4 il faut se rappeler que les valeurs de probabilité interprètent des degrés de croyance. Il faut aussi remarquer que P_s4 pose en réalité une condition de cohérence sur les degrés de croyance de l'agent. Ainsi, la réponse à la première question semble plutôt négative alors que la réponse à la seconde question est positive. Par conséquent, P_s4 doit être considéré comme un axiome de rationalité. Cependant, il est indéniable que la propriété qu'il énonce est essentielle pour que les degrés de croyance s'interprètent comme des probabilités. Restent les axiomes P_s5, P_s6, P_s7 qui expriment des propriétés formelles nécessaires à la démonstration des théorèmes ; ils sont considérés comme des axiomes de structure. Il n'y a pas de motivation indépendante qui indiquerait pourquoi un agent devrait les respecter. Cependant, les axiomes de structure ont une importance dans l'interprétation de la théorie. Dans la mesure où ils sont requis pour la

démonstration du théorème de représentation, ils expriment *aussi* des contraintes pour l'agent. Cependant, ces contraintes ne sont pas des contraintes de cohérence, elles ne sont pas motivées par des intuitions substantielles concernant la rationalité des préférences et des choix. En ce sens, ils correspondent peut-être à une part d'arbitraire dans la théorie et certains auteurs, comme Shaffer et Joyce, considèrent que pour cette raison, la plupart des théorèmes de représentation que nous connaissons, pour les théories standard, ne démontrent pas véritablement l'adéquation des théories de la rationalité qu'ils sont censés valider. Pour atténuer la difficulté que semble poser l'interprétation des axiomes de structure, nous souscrivons à la solution mise de l'avant par Jeffrey et Joyce qui ont recours à une *condition d'extensibilité* de la structure de préférence d'un agent[34]. Dans cette interprétation, on ne prétend pas que l'ordonnancement des préférences d'un agent rationnel doive obéir aux axiomes de structure, mais plutôt qu'il doit exister une extension cohérente de la structure de préférence d'un agent qui obéit aux axiomes de structure sans contredire aucun des axiomes de rationalité. Nous aurons l'occasion de revenir sur cette interprétation de la théorie de l'utilité espérée à l'occasion de notre discussion de la délibération chez Jeffrey et de l'usage du lemme de Lindenbaum dans ce contexte. Notons pour l'instant que cette interprétation, qui nous semble une solution très attrayante pour l'utilitarisme bayésien axiomatisé, entraîne une difficulté du côté du théorème de représentation de Savage. En effet, il est clair qu'il existera, en général, plus d'une façon « d'étendre » une structure de préférence pour la compléter. Or, le théorème de représentation pour le système de postulats de la théorie de Savage doit établir que les fonctions u et P sont uniques modulo le choix d'un point zéro et d'une unité de mesure —

autrement dit, il doit établir que u et P sont unique modulo les transformations linéaires. Chaque extension cohérente d'une structure de préférence possédant sa propre fonction d'utilité et de probabilité. Par conséquent, l'interprétation des axiomes de structure par la *condition d'extensibilité* nous fait perdre un qualificatif important du théorème de représentation, l'unicité des fonctions u et P.

Notre second commentaire porte sur la définition des actes dans la théorie de Savage. À la suite de Joyce[35], on peut exprimer le concept d'acte selon Savage comme un axiome « caché » de la théorie. Pour ce faire, il faut formuler les notions d'acte constant, d'acte conditionnel et d'acte mixte d'une façon perspicace. Le résultat est une définition de $\boldsymbol{A}$ qui permet de mettre en lumière la structure particulière de l'ensemble des actes chez Savage.

(Richesse de $\boldsymbol{A}$) Pour tout résultat $c \in \boldsymbol{C}$, $\boldsymbol{A}$ contient l'acte a_0 qui est défini par

$a_0 = c$ pour chaque état s dans $\boldsymbol{S}$; (c.-à-d. a_0 est constant) et

a_0 est un pari tel que si s est réalisé alors c.

De plus, pour tous les actes a et b, un événement quelconque E et les résultats c et c' définis par les clauses $a(s) = c$ et $b(s) = c'$, pour chaque état $s \in S$, $\boldsymbol{A}$ contient l'acte mixte

$a_E \,\&\, b_{\sim E} =$ le pari tel que $(s \;\&\; s \in E \Rightarrow c) \,\&\, (s \,\&\, s \in \sim E \Rightarrow c')$ où

$(s \;\&\; s \in E \Rightarrow c)$ dénote l'acte conditionnel sur E ayant pour conséquence c [36].

Tel est l'ensemble $\boldsymbol{A}$ qui est vraiment requis pour la théorie de Savage. Comme le note Joyce, Savage n'a pas besoin de tous les actes que contiendrait l'ensemble $\boldsymbol{A}$ — tel que Savage définit $\boldsymbol{A}$ — et qui contient *toutes* les fonctions de $\boldsymbol{S}$ dans $\boldsymbol{C}$. Il lui suffit d'avoir les actes constants et la condition de fermeture de $\boldsymbol{A}$ sous les actes mixtes. Cependant, il appert que la taille et la composition

de l'ensemble des actes de Savage donne lieu à des critiques. Nous avons vu que le postulat P_s6 impliquait que S était de cardinalité indénombrable. Le vaste ensemble 2^S semble à la fois trop large et trop peu caractérisé pour correspondre à l'idée intuitive d'action intentionnelle, ou simplement à l'idée d'un agent qui *fait en sorte que* quelque chose se produise. Il y a donc un problème d'adéquation matérielle à propos du concept d'acte de Savage. Comme le note Joyce, aucun être humain ne peut accomplir un acte constant car personne ne peut tout prévoir[37]. De plus, le concept d'acte de Savage donne lieu à un problème d'adéquation formelle que nous devons mentionner. En effet, comme l'économiste et mathématicien Peter Wakker [1993] l'a montré, le respect des axiomes de Savage n'implique pas que le principe de dominance stochastique, P_s4 dans notre formulation, soit respecté par les agents réels. Au contraire, il semble qu'on trouve aisément des contre-exemples. Selon Wakker, il y a violation dès que le co-domaine de la fonction d'utilité est suffisamment riche, par exemple, s'il contient un intervalle et que la fonction de mesure est constructive. En particulier, dans un tel cas, si tous les axiomes de Savage sont respectés et que la condition de monotonicité stricte[38] de la relation de préférence sur les états est violée, un agent accepterait d'échanger un acte pour un autre qui donnerait un résultat strictement moins bon. On peut construire une pompe à fric contre un tel agent. Toujours selon Wakker, la dominance stochastique et la monotonicité sur les états sont deux conditions qui sont toujours satisfaites lorsqu'on restreint le domaine des actes aux actes qui ont au plus des résultats en quantité finie[39].

Il n'y a pas que la taille de l'ensemble des actes qui a fait l'objet de critiques, mais aussi le statut particulier des actes imaginaires et des actes

constants. Considérons d'abord la question des actes imaginaires. Ce sont des actes à propos desquels on peut délibérer, mais qui ne sont pas « concrets » au sens où l'on ne peut pas les réaliser ; ils ne sont pas faisables. Glenn Shafer, dont l'article «Savage Revisited» comprend une discussion critique approfondie des postulats de Savage, apprécie de façon nuancée l'inclusion des actes imaginaires[40]. Il note que plusieurs des auteurs qui ont proposé des critiques de la théorie Savage, dont Fishburn que nous avons abondamment cité, refusent de concevoir un agent qui aurait une structure de préférence complète pour les actes imaginaires. Comme Shafer, nous croyons que Savage a raison de penser que l'on peut délibérer à propos de *certains* de ces schèmes de conséquences (*patterns of consequences*) même s'ils ne sont pas faisables. Par exemple, un adolescent peut bien avoir des préférences bien définies sur l'intérêt d'acquérir tel ou tel modèle de voiture sport qu'il ne peut pas s'offrir ou qu'il ne pourra jamais acheter dans un avenir prévisible. Ainsi, la critique des actes imaginaires a moins de poids que la critique des actes constants qui posent également des problèmes d'interprétation.

Les actes constants ont été définis comme des actes dont le résultat est constant, de façon indépendante des états du monde. Le problème des actes constants est qu'il est facile, pour la plupart des actes, d'imaginer des circonstances possibles où le résultat espéré ne se produira pas. L'acte de gratter une allumette ne produira pas de feu en l'absence d'oxygène, si l'allumette est mouillée, et ainsi de suite pour une foule d'actes ordinaires. Pour la plupart des actes envisageables, il y a donc certains états du monde où ils ne produiront pas l'effet attendu. Comme disent les théoriciens de la décision, cette difficulté constitue un *problème d'applicabilité* de la théorie, ou pour reprendre l'expression que nous utilisons un problème d'adéquation

matérielle. Luce et Krantz [1971] ont réussi à formuler une théorie comparable à celle de Savage qui élimine le recours aux actes constants. Le système de postulats de ces auteurs, qui définissent les relations de préférence sur des fonctions partielles, est cependant critiqué lui aussi car il admet un ensemble d'actes trop large, avec la possibilité d'avoir des préférences définies pour des actes qui pourraient s'avérer impossibles. Pour cette raison, il n'est plus à l'ordre du jour d'essayer de justifier indûment les actes constants. Selon Joyce, la source du problème est la taille indue de l'ensemble des actes, la totalité de 2^S, qui doit contenir absolument toutes les fonctions de l'ensemble des états dans l'ensemble des résultats[41]. Comme nous l'avons observé dans l'exemple des actes constants, plusieurs de ces fonctions ne correspondent pas à des actes concevables ou des situations possibles.

Notre commentaire suivant porte sur le contre-exemple proposé par Maurice Allais dans ses fameux articles Allais [1953] et Allais [1979]. Ce contre-exemple concerne le principe de la chose sûre et l'axiome d'indépendance. Il appartient à la même classe de difficultés que le non-respect de la transitivité des préférences que nous avons déjà discuté au chapitre II. Comme nous le disions, l'échec de la transitivité des préférences et le non-respect du principe de la chose sûre plaident contre une interprétation descriptive de la théorie de l'utilité espérée. Mais pour ceux qui favorisent une interprétation normative, et au premier titre, pour Savage et Jeffrey, ces difficultés ne sont pas reçues comme des objections fortes[42]. Néanmoins, nous croyons que ces difficultés mettent en lumière un phénomène de sensibilité au contexte qui devrait être pris en compte dans une analyse de la délibération.

Le problème posé par Allais concerne le comportement d'un agent rationnel face au risque et révèle une survalorisation d'un gain assuré

comparativement à un choix risqué qui serait plus avantageux selon le calcul de l'utilité espérée. On propose à un agent de prendre deux décisions dans le but de vérifier la cohérence des préférences que vont révéler ses choix :

(1) Préférez-vous la situation A ou La situation B ?

Situation A :

certitude de recevoir 100 millions de dollars.

Situation B :

10% de chance de recevoir 500 millions de dollars ;

89% de chance de recevoir 100 millions de dollars ;

1% de chance de ne recevoir rien du tout..

(2) Préférez-vous la situation C ou la situation D

Situation C ;

11% de chance de recevoir 100 millions de dollars ;

89% de chance de ne recevoir rien du tout.

Situation D ;

10% de chance de recevoir 100 millions de dollars ;

90% de chance de ne recevoir rien du tout.

Plusieurs personnes prudentes et réfléchies vont préférer (A) (la chose sûre) à (B) dans le premier problème de choix et (D) à (C) dans le second problème de choix. Pourtant, si le postulat d'indépendance (P_s2) est respecté, et qu'on applique la règle du calcul de l'utilité espérée, un agent qui juge que $B \prec A$ devrait juger que $D \prec C$.

En effet, on établit les valeurs suivantes pour les diverses options :

u(A)= u(100 M$)

u(B)= 0.1 u(500 M$) + 0.89 u(100 M$) + 0.01 u(0 $)

u(C)= 0.11 u(100 M$) + 0.89 u(0 $)

u(D)= 0.1 u(100 M$) + 0.90 u(0 $).

Les lois de l'arithmétique font que

u(A) – [43]u(B) = 0.11 u(100 M$) – [(0.1 u(500 M$) + 0.01 u(0 $)]

u(C) – u(D) = 0.11 u(100 M$) – [(0.1 u(500 M$) + 0.01 u(0 $)]

autrement dit, la différence de valeur entre d'une part, A et B et d'autre part, C et D est égale. Il y a donc une conséquence commune dans les deux situations de choix et le principe de la chose sûre est violé.

Il y a plusieurs façons de répondre au paradoxe de Allais et nous allons nous limiter à discuter celle de Savage lui-même » Voici deux citations qui proviennent à quelques lignes près, du même passage :

> Plusieurs exceptions apparentes à la théorie s'avèrent ne pas être des objections du tout.
>
> Si, après une délibération approfondie, quelqu'un peut tenir une paire de préférences distinctes qui est en conflit avec le principe de la chose sûre, il doit abandonner ou modifier le principe. [...] Une personne qui a accepté une théorie normative à titre provisoire doit étudier consciencieusement les situations où la théorie semble l'induire en erreur et décider si, après réflexion, il conserve sa première impression de la situation ou s'il accepte les conséquences de la théorie pour cette situation. [44]

Comme on peut le constater, Savage ne recourt pas à l'interprétation normative comme à une sorte d'hypothèse *ad hoc* pour sauver la théorie. Une théorie normative n'est pas entièrement immunisée contre les résultats

expérimentaux. Cependant, des recherches ont montré que les agents qui choisissaient à l'encontre de la théorie n'avait pas des explications raisonnables ou plausibles pour motiver leurs choix[45]. Dans la situation du problème posé par Allais, Savage rapporte que sa première réaction a été d'adopter des préférences contraires au calcul de l'utilité, (autrement dit B $\prec$ A et C $\prec$ D) de maintenir ces préférences après réflexion et de les trouver intuitivement motivés. Mais pour analyser ses préférences, il a modifié la situation en fusionnant les options proposées par Allais dans une loterie unique possédant 100 tickets distincts[46]. Sous cette reformulation, Savage découvre que ses préférences sont conformes à la théorie de l'utilité (autrement dit B $\prec$ A et D $\prec$ C).

Notre commentaire suivant concerne l'assimilation du risque et de l'incertitude. Comme la théorie de Ramsey, la théorie de Savage identifie entièrement l'incertitude de l'agent au risque qu'il prend en faisant un choix. Cette assimilation de l'incertitude au risque fait l'objet d'une critique bien connue due à Daniel Ellsberg qui propose de distinguer ces deux concepts[47]. On peut parler de risque lorsque les probabilités sont connues ou que les valeurs de probabilité peuvent être estimées mais on doit parler d'incertitude ou d'ignorance lorsqu'il n'est pas possible de donner une valeur de probabilité qui ne soit pas arbitraire. Ainsi, si je dois prendre une boule dans une urne qui contient 100 boules parmi lesquelles il y a des boules rouges et des boules noires dans une proportion inconnue, je ne peux pas appliquer la règle de décision utilitariste car je n'ai pas de valeur pour estimer la probabilité d'une boule rouge. Ellsberg présente son problème comme un contre-exemple à la valeur descriptive d'adéquation du principe d'indépendance de Savage (P_s2) et non comme un problème de légitimité normative. En ce sens on peut le

considérer comme une difficulté qui s'apparente au problème de Allais que nous venons de discuter. Comme le montre Resnik [1987], on peut construire une réponse au paradoxe de Ellsberg qui soit calquée sur la réponse de Savage pour le problème de Allais[48]. Autrement dit, on peut esquiver ce paradoxe en l'analysant comme une incomplétude des préférences[49]. Il nous semble cependant que du point de vue de la délibération, la distinction entre le risque et l'ambiguïté est significative et qu'elle devrait recevoir l'attention qu'elle mérite.

Les commentaires qui précèdent ont signalé quelques critiques importantes à l'endroit de la théorie de Savage. En présentant le modèle de la délibération de la théorie de Savage dans la section suivante, nous aurons l'occasion de mentionner d'autres difficultés dont le problème des petits mondes (*small worlds*).

Comme nous l'avons vu, les postulats de Savage ne doivent pas être interprétés comme une description, ni comme une méthode de prise de décision. De même, on pourrait dire que le fait de maximiser l'utilité n'est pas un but poursuivi, autrement dit, un objectif de l'agent. La question globale se pose : qu'est-ce qui a été accompli par l'explicitation de la théorie de l'utilité de Savage ? L'interprétation que nous avons adoptée est la même pour toutes les logiques de la décision qui s'apparentent au modèle standard. La logique de la décision de Savage est une description d'une norme de rationalité, une description de ce que cette norme exige des choix d'un agent. Les postulats de la théorie sont des contraintes qui restreignent la classe des choix possibles en tentant de cerner la classe des bons choix. C'est dans cette perspective qu'on peut offrir une réponse à la question globale. Ce qui est accompli du point de vue logico-philosophique, c'est l'élucidation d'un concept de « valeur d'un

choix » (*choiceworthiness*) pour un agent individuel qui est confronté au choix dans l'incertitude.

4.2.4 Le modèle de la délibération chez Savage

Nous pouvons maintenant tenter de rendre explicite le modèle de la délibération qui est contenu implicitement dans la théorie que nous venons d'exposer. Pour y parvenir, notre point de départ est le fameux problème de l'omelette que Savage a rendu célèbre. Ce problème permet d'illustrer les éléments d'analyse d'une situation de choix dans la théorie de Savage. Nous pourrons constater que la théorie de la décision de Savage, comme toutes les théories standards, contourne assez aisément les difficultés qui sont normalement associées à la délibération, par exemple les phénomènes liés à la prise en compte de la temporalité, de la planification de l'action, de l'intentionnalité ou de la faiblesse de la volonté. Ce faisant nous allons mettre en évidence des aspects importants du mécanisme de délibération en logique de la décision. La *situation* de ce problème de choix est celle d'un agent qui prépare une omelette et qui a déjà cassé cinq œufs dans un bol. Avant de casser le sixième œuf, il *délibère* à propos d'actions faisables ou possibles alors qu'il tient en main le sixième œuf qui est soit bon, soit pourri. Selon l'état du sixième œuf et l'acte pour lequel il va opter, les conséquences seront plus ou moins avantageuses. Le problème de l'omelette est représenté dans un tableau de décision qui indique les actions possibles, les états du monde et les résultats associés à chaque couple {action, état}. Savage note que ce petit exemple illustre bien la variété des résultats qui peuvent importer pour un agent dans une situation de choix qui implique une incertitude. On remarque

que les préférences de l'agent ne sont pas exprimées par des enjeux monétaires. Néanmoins, le choix de l'agent se laisse aisément concevoir comme un pari. L'action la plus prudente est de casser l'œuf dans une soucoupe pourvu que l'inconvénient de laver une soucoupe soit jugé moindre que celui d'avoir une omelette à 5 œufs. Cette solution se justifie directement par les principes d'admissibilité et de dominance.

Actes	*États*	
	bon	pourri
casser et ajouter au bol	omelette à 6 oeufs	pas d'omelette, 5 œufs perdus
casser dans une soucoupe	omelette à 6 œufs, soucoupe à laver	omelette à 5 œufs, soucoupe à laver
jeter l'œuf	omelette à 5 œufs, un bon œuf perdu	omelette à 5 oeufs

On peut constater que dans la théorie de Savage, les actes sont vraiment identifiés à leurs conséquences : deux actions qui auraient les mêmes conséquences dans tous les états du monde sont identiques[50]. Par exemple, il n'y a pas lieu de distinguer dans ce problème l'acte qui consiste à casser un œuf dans une soucoupe, de le casser dans un petit bol différent ou dans une tasse. Au contraire, les actions distinguées dans le tableau engendrent des résultats différents pour au moins un état du monde. Ainsi, bien que dans une description formelle d'une situation de choix, on pourrait vouloir remplacer le premier acte par « briser dans un bol et si l'omelette est gâchée, aller acheter

des croissants » et plusieurs autres actes imaginables d'un même type, la famille de tous les tableaux possibles contenant de tels actes est contenue comme une toile de fond dans le tableau que présenté. Dans la théorie de Savage, on dit que la situation pour l'agent prend place dans un *petit monde* sur la toile de fond d'un *grand monde* qu'il serait « ridicule » d'essayer de décrire. Il n'y a pas de critère précis qui permette de délimiter un petit monde[51]. Ce qu'on appelle le « problème des petits mondes » tient au fait qu'une analyse de la meilleure option d'un problème de décision qui vaut dans un petit monde pourrait recevoir une solution différente pour un autre petit monde qui serait plus détaillé. C'est le sophisme que les environnementalistes ne cessent de dénoncer ; une décision peut être rationnelle en fonction de conséquences que nous savons apprécier dans un contexte spatio-temporel limité mais devenir absurde lorsqu'on prend en considération le portrait global de la situation. Savage nous met également en garde contre la tentation de surestimer la précision avec laquelle un agent peut se représenter les résultats à propos desquels il délibère,

> En fin de compte, une conséquence est une idéalisation dont on ne peut possiblement jamais faire une approximation juste.[52]

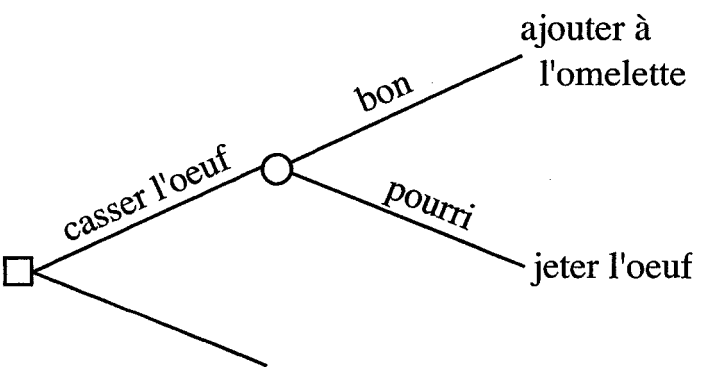
Au chapitre VI, nous aurons l'occasion d'exposer la structure arborescente de la temporalité qui éclaire la question fort importante de savoir comment un petit monde est relié à un autre. La distinction entre le « petit monde » et le « grand monde » ne doit pas nécessairement être perçue comme une difficulté — interne ou externe — de la théorie. De façon très éloquente, Picavet exprime le réalisme et la perspicacité de Savage.

> Savage, en un aperçu admirable, rend ainsi sensible le caractère tout relatif du langage des moyens et des fins dans la description de l'action humaine. Nous ajustons l'action aux résultats que nous en attendons, de sorte que l'élection des moyens et des stratégies — l'action même, donc — est au fond une certaine façon de se préparer à ce qui doit arriver. Mais les résultats auxquels il faut s'attendre sont inévitablement aussi des paris sur l'avenir, de sorte que les fins que l'on visait apparaissent elles-mêmes comme des actes préparant le futur.[53]

Il y a là un bon argument polémique à l'endroit d'une certaine critique philosophique de la théorie de la décision qui voudrait en réduire la signifiance sous prétexte qu'elle n'est qu'instrumentale. Ce que nous dit Picavet, c'est que la capacité prédictive des humains étant limitée à l'horizon des petits mondes, toutes les finalités humaines, peut-être même le contentement et le bonheur, peuvent être appelées à devenir des moyens en vue de finalités plus grandes qui les englobent. On peut aussi illustrer le problème de la surabondance des actes que nous avons déjà mentionnés. Si l'ensemble des options qui s'offrent à l'agent devait contenir toutes les fonctions de $\boldsymbol{S}$ dans $\boldsymbol{C}$, il devrait contenir aussi l'acte d'ajouter l'œuf pourri aux cinq autres et d'obtenir une omelette de six œufs ainsi que l'acte qui consiste à ajouter l'œuf pourri aux cinq autres et de perdre les cinq œufs. Ces actes semblent fort étranges parce qu'ils heurtent nos croyances concernant les œufs et les omelettes, mais néanmoins ils font partie des options que doit considérer un agent. Ce sont les premières difficultés de la délibération dans la théorie de Savage. Le problème de la taille de la classe des actes a été observé depuis longtemps par Fishburn et il est formulé de façon dramatique par Jeffrey dans le passage suivant.

> Dans le système de Savage, les options (*prospects*) semblent être des entités entre lesquelles un être aux capacités finies ne pourrait pas choisir. Pour Savage, le choix d'un « acte » est le choix d'un schème d'action qui associe une conséquence définie pour chacun des états parmi une infinité d'états de la nature. Si tels sont les actes, alors seul Dieu pourrait savoir quels sont les actes qui sont performés. Après avoir fait un acte, un agent humain pourrait apprendre les conséquences de son acte associé à l'état de la nature qui prévaut, mais ni avant, ni après la performance de son acte ne pourrait-il connaître les conséquences que son acte associe à chaque état de la nature.[54]

Nos deux observations suivantes portent respectivement sur la composition et la réduction des actes. Dans le tableau qui décrit le problème de l'omelette, on pourrait croire que les actes possibles sont réellement des actes composés. En effet, l'acte de casser l'œuf dans une soucoupe est vraiment un acte conditionnel de la forme « je casse l'œuf et s'il est bon, je l'ajoute à l'omelette etc. ». La question qui se pose est celle de la composition des actes et elle se pose dans toutes les théories de la décision. Ici, dans le contexte de la théorie de Savage et du problème de l'omelette, on peut représenter la situation par un arbre de décision :



Dans ce diagramme, deux sortes de nœuds sont distingués, un nœud avec un carré représente un point de choix (c'est l'agent qui choisit) tandis qu'un nœud

avec un cercle représente un point de chance (c'est « la nature » qui choisit[55]). Dans l'analyse de Savage, il y a un principe de fusion interne des actes — on pourrait dire une fusion des *gestes* qui font l'acte — qui fait de l'acte casser l'œuf et l'ajouter à l'omelette un acte unique, distingué de l'acte de casser l'œuf et de le jeter. Ce principe que nous appelons la fusion interne des gestes qui constituent l'acte est une autre caractéristique du modèle standard qui assimile l'action et son résultat.

La question de la réduction des actes concerne le concept de choix séquentiel dans un contexte où l'on admet la réduction des loteries composées. On dit d'une loterie qu'elle est composée lorsqu'elle correspond à un pari conditionnel comparable à ceux que nous avons examinés en exposant la théorie de Ramsey au chapitre II. Considérons par exemple le pari suivant : « Si la France remporte son match de semi-finale, je parie qu'elle gagnera face au Brésil en finale. » Le principe de réduction des loteries composées nous dit qu'à chaque fois que l'on peut décrire un pari conditionnel de la forme « si la France gagne, alors *je vais* parier 10$ qu'elle gagnera face au Brésil », je peux éliminer la temporalité de cette description de l'acte conditionnel en considérant un pari sur une conditionnelle que je fais dès maintenant et qui dit « je parie 10$ que si la France gagne alors elle vaincra en finale ». Si on se rappelle que, dans la foulée de Ramsey, toute action équivaut à un pari, on peut comprendre toute la portée de cette réduction. En généralisant cette équivalence, on obtient le principe de réduction des loteries composées. Exprimé de façon informelle, il stipule que tout pari conditionnel est indifférent (au sens exact introduit pour la relation ...≈...) à un pari simple sur les conséquences $c_1,\ldots, c_n$, conséquences pour lesquelles les probabilités sont calculées en respectant les axiomes de Kolmogorov. En théorie des jeux,

ce principe de réduction est utilisé systématiquement pour passer de la représentation de la forme extensive d'un jeu (arborescence) à la forme stratégique (tableau). Il faut remarquer que c'est Savage lui-même qui parle de la réduction des actes conditionnels comme une méthode pour éliminer la temporalité de l'analyse de ce genre de situations[56]. On peut remarquer que si la temporalité était réellement prise en compte, le principe de réduction des loteries composées ne serait pas valide. Il se pourrait, par exemple, que je n'aie pas les dix dollars maintenant pour débourser la mise de ce pari. Ainsi, pour moi, il ne serait pas indifférent de parier maintenant ou de parier plus tard. La perspective de voir toutes les décisions de la vie d'une personne amalgamées dans un grand plan qui ne comporterait qu'un seul choix mesure la difficulté de réconcilier le principe de réduction des loteries composées avec nos intuitions concernant la délibération réelle des agents. Savage la qualifie simplement « d'irréaliste »[57].

Nous complétons cette discussion en mentionnant une double difficulté que la théorie de Jeffrey va tenter de corriger. Dans la théorie de Savage, la valeur d'une action est déterminée par des valeurs d'utilité des conséquences qui sont inconditionnelles, autrement dit, qui ne sont pas conditionnelles aux états du monde. Comme chez Ramsey, l'utilité est déterminée en fonction de biens considérés comme ultimes. Dans le problème de l'omelette, on pourrait trouver plus naturel de répartir les valeurs d'utilité directement sur les résultats (omelette à 6 œufs, omelette à 5 œufs, pas d'omelette). Comme nous l'avons vu, l'analyse de Savage combine l'inconvénient avec le résultat dans les cases qui sont les valeurs des couples {action, état}[58].

De même, toujours dans la théorie de Savage, la valeur d'une action est déterminée par des probabilités d'états de manière non-conditionnelle. On

traite les états comme étant indépendants des actes du point de vue des probabilités. Ceci introduit une limitation dans l'application de la théorie, il est raisonnable de chercher à maximiser l'utilité espérée seulement si les degrés de croyance de l'agent à propos de la probabilité des divers états du monde ne dépendent pas des actes qu'il performe. Ce phénomène est suffisamment important dans l'appréciation de la théorie de Savage pour que nous prenions le temps de l'illustrer.

Un bon exemple de la dépendance des actes aux états nous est proposé par Joyce[59]. Imaginons que nous venons de garer la voiture dans un quartier louche et qu'un garçon nous approche pour nous proposer de surveiller et de protéger notre voiture pour la somme de 10$. On vous a dit que les gens qui refusaient de payer pour la protection se retrouvaient invariablement avec un pare-brise éclaté alors que ceux qui acceptaient de payer retrouvaient leur voiture intacte. Vous voyez bien qu'il s'agit d'extorsion, mais qu'allez-vous faire ? Le remplacement de votre pare-brise coûte 400$ et vous n'avez aucune autre option pour garer votre voiture et vous rendre à la réunion pour laquelle vous êtes déjà en retard. Les postulats P_s2 et P_s3 impliquent que vous ne devriez pas payer les 10$ de protection car il vaut mieux avoir ces 10$ dans vos poches, que le pare-choc soit brisé ou non. Dans ce cas, l'option de refuser de payer a vraiment un effet sur la probabilité de l'état qui correspond au pare-brise éclaté. Savage dirait que les postulats de la théorie ont été appliqués à un problème qui est mal posé. Comme le remarque Joyce, tout dépend de la partition des événements dans la formulation du problème. Si nous remplaçons la partition implicite du problème précédent

E : Le pare-brise est brisé.

$\sim E$: Le pare-brise n'est pas brisé.

par la partition

E_1 : Le pare-brise est brisé, quoi que tu fasse.

E_2 : Le pare-brise n'est pas brisé si tu paies et il est brisé si tu ne paies pas.

E_3 : Le pare-brise est brisé si tu paies et il n'est pas brisé si tu ne paies pas.

E_4 : Le pare-brise n'est pas brisé, quoi que tu fasses.

où l'acte E_2 est relativement plus probable que les autres événements, le problème est dissous et l'acte dominant est celui qui produit E_2, c'est-à-dire l'acte de payer le montant de 10$ [60].

La théorie de Jeffrey veut apporter une solution à cette « difficulté » et voudrait éliminer en partie cette dépendance à la partition des événements.

4.3 La logique de la décision de Richard Jeffrey

En philosophie, la logique de la décision de Jeffrey est considérée à bon droit comme la référence principale en logique de la décision. À l'extérieur du domaine de la philosophie, en économie ou en recherches opérationnelles par exemple, la théorie de Jeffrey est moins connue. Il y a néanmoins un consensus sur un point : cette logique surpasse sans contredit toutes celles que nous avons étudiées jusqu'ici et inspire le respect à tous ceux qui l'ont fréquenté, même à ses détracteurs. Ainsi, comme le note James Joyce après David Lewis et Brian Skyrms, même lorsque nous devrons conclure qu'elle n'est pas entièrement satisfaisante, nous aurons observé qu'aucune théorie ne saurait la surpasser sans reprendre plusieurs de ses caractéristiques. Comment

résumer en quelques mots l'objectif de Jeffrey dans *The Logic of Decision* ? Le voici en ses propres termes

> [...] rappelons le terme central de ce livre : En un certain sens, l'explication bayésienne de la délibération fournit une logique de la décision.[61]

Il note cependant dans le même passage que le développement de ce thème central ne cesse de nous conduire sur des avenues périphériques. Richard Jeffrey a eu pour professeurs deux grandes figures de l'empirisme logique, Rudolf Carnap à Chicago (1946-51) et Carl G. Hempel à Princeton (1955-57). À la différence de Savage qui était mathématicien, Jeffrey est un philosophe dont les préoccupations dépassent les horizons de la logique et de la théorie de la décision pour inclure la problématique des fondements des probabilités et surtout l'épistémologie des jugements probabilistes, sujet sur lequel il reviendra jusque dans son dernier ouvrage publié, *Subjective Probability : The Real Thing* [62].

La logique de la décision de Jeffrey est d'abord parue dans la première édition de son *The Logic of Decision* en 1965. Le fait que cet ouvrage se présente comme une introduction peut entraîner un malentendu sur la nature et l'originalité de son contenu. Pour mener à bien l'élaboration de sa théorie, il a pu compter sur la collaboration de nombreux philosophes et mathématiciens parmi lesquels on compte, en plus de Carnap et Hempel qu'on a déjà nommés, Abner Shimony, Patrick Suppes, Kurt Gödel et surtout Ethan Bolker qui a démontré les théorèmes sur « les fonctions ressemblant à des quotients de mesures » qui permettent d'établir un théorème de représentation qui est original et spécifique à la théorie des préférences de Jeffrey[63].

La théorie de la décision est revenue sous une forme révisée dans la deuxième édition de 1983 et corrigée sur un point d'importance en 1990. Pour bien comprendre la logique des probabilités que Jeffrey a élaborée sur une période qui couvre près d'un demi-siècle, il faut tenir compte de la somme impressionnante d'articles qu'il a publiés sur le sujet dans les quarante dernières années. Quelques-unes de ses plus importantes contributions ont été publiées dans un recueil, Jeffrey [1992]. De plus il est nécessaire de comprendre les principales théories qui forment l'arrière-plan des débats récents pour cerner ce qu'il y a d'original et de spécifique dans la position que défend Jeffrey en regard de théories antérieures, comme celles de Ramsey et Savage ou en regard de théories rivales ou plus récentes comme la théorie causale de la décision[64].

4.3.1 Le calcul de la désirabilité

Comme nous l'avons dit dans la section précédente, la théorie de Jeffrey est une généralisation de celles de Savage et Ramsey. Nous disions que les éléments de base de l'analyse d'un problème de décision sont unifiés sous le concept de proposition dans la théorie de Jeffrey. Les options (*prospects*) sont des propositions et la relation de préférence peut aussi bien s'appliquer aux actes qu'aux événements que Savage distinguait. Nous disions aussi que la dépendance des actes et des états est prise en compte dans la forme du calcul de la désirabilité d'une action chez Jeffrey. On peut illustrer cette différence en comparant les deux règles. Comme dans l'énoncé du théorème T_s2 de la section précédente, on a E_i, pour $i = 1,\ldots, n$, une partition de S, un acte a

ayant pour conséquence c_i et une fonction d'utilité u, la valeur espérée de l'acte a notée $U(a)$

$$U(a) = \sum_i u(c_i)\, P(E_i)$$

Voyons maintenant la règle analogue chez Jeffrey. Nous expliquerons plus loin la construction de l'algèbre de Boole des propositions qui sont les porteurs des valeurs de probabilité. Pour l'instant, considérons l'ensemble Ω de tous les mondes possibles (de tous les états possibles de la nature) et notons le $\top$. La probabilité inconditionnelle est notée $P(\ldots)$ et la désirabilité est notée $des(\ldots)$. Si les états S_i forme une partition de $\top$, A est un acte et $P(X \mid A)$ est la probabilité, dite conditionnelle, de X étant donné A, on a

$$des(A) = \sum_i des(S_i \wedge A)\ P(S_i \mid A)$$

Nous allons développer notre examen de la théorie de Jeffrey par une explication des termes de cette formule. Si on considère cette formule d'un point de vue purement formel, elle énonce une propriété abstraite qui décrit la combinaison de la désirabilité et de la probabilité. Plus précisément, elle exprime l'idée que la relation principale entre *des* et *P* est que leur produit est additif, tout comme *P* elle-même est additive[65]. Dire que la probabilité est additive, c'est dire que la probabilité d'une proposition est la somme des probabilités pour chaque circonstance où elle est vraie. En comparant les deux formules, on note que la désirabilité d'un acte pour Jeffrey est l'utilité espérée au sens de Savage de l'acte conditionnalisé sur lui-même, c'est-à-dire,

conditionnel à sa propre réalisation. Jeffrey explique qu'il a choisi le terme « désirabilité » pour éviter les connotations du terme « utilité » qui renvoient à l'hédonisme de Bentham. La désirabilité est l'utilité espérée conditionnelle, elle est aussi subjective parce qu'elle se compose des préférences de l'agent et de valeurs de probabilité qui sont des degrés de croyance. On mesure la désirabilité d'une proposition en calculant la moyenne pondérée de la désirabilité des cas où cette proposition est vraie, où le poids de chacun des cas est proportionnel à sa probabilité. Jeffrey exprime le calcul de la désirabilité par un axiome.

(Axiome pour *des*) Soit A et B, deux propositions distinctes qui représentent des options mutuellement exclusives, c.-à-d., P (A & B) = 0 ; on pose que P (A $\vee$ B) $\neq 0$

$$des(A \vee B) = \frac{P(A)des(A) + P(B)des(B)}{P(A) + P(B)}$$

On peut vérifier la valeur intuitive de cette formule en examinant un exemple où il est raisonnable d'évaluer la désirabilité d'une disjonction. Mon employeur va défrayer le coût de mon voyage à Toronto, mais je ne sais pas quel sera le moyen de transport qu'il va choisir pour mon déplacement. Selon les coûts associés à ces options, je peux estimer la probabilité qu'il opte pour l'un ou l'autre moyen de transport

	temps	probabilité
voiture	7 heures	0,5
avion	2 heures	0,2
train	5 heures	0,3

La désirabilité d'un moyen de transport est une fonction inverse du temps requis pour faire le voyage. Ainsi, la désirabilité de faire le voyage en voiture est de 1/7, soit 0,14 et l'on complète le tableau de la même façon pour l'avion et le train :

voiture	0,14
avion	0,5
train	0,2

À l'aide de ces tableaux, je peux calculer la désirabilité d'aller à Toronto en avion (*A*) ou en train (*T*), soit

$$des(A \vee T) = \frac{P(A)des(A) + P(T)des(T)}{P(A) + P(T)}$$

$$= \frac{(0,2 \times 0,5) + (0,3 \times 0,2)}{0,5 + 0,2} = \frac{1,6}{0,7} = 2,29$$

Cet exemple de calcul illustre la nature de la délibération pour la désirabilité d'une alternative, ce qui est le cas le plus général. On note que le problème de la calibration de l'échelle de préférence, problème qui donnait lieu aux

propositions éthiquement neutres dans la théorie de Ramsey, ne se pose pas ici. Les préférences se comparent en termes de temps de déplacement, mais on pourrait prendre le trajet le plus court comme unité de référence et rien ne serait changé.

4.3.2 Les préférences

Pour caractériser davantage la théorie de Jeffrey, il nous faudra examiner l'idée de probabilité conditionnelle, ce que nous ferons un peu plus loin, mais nous allons d'abord nous attarder à l'analyse des préférences. Cette logique de la préférence est au centre de la théorie de Jeffrey.

Il est clair que les préférences ne s'appliquent pas seulement qu'à des options dans un problème de choix; on peut en avoir pour une journée de demain qui soit ensoleillée plutôt qu'un lendemain pluvieux. Il est facile de rendre compte de telles préférences en prenant les propositions comme objet de préférence : on peut avoir une préférence pour la vérité de la proposition *qu'il fera beau demain.*

Ce concept de probabilité a une logique assez simple puisqu'elle est gouvernée par deux lois évidentes, la transitivité et la trichotomie

$$(A \succ B, \text{ ou } A \approx B, \text{ ou } B \succ A)$$

La théorie formalisée par Bolker et Jeffrey comprend aussi deux règles qui sont utilisées dans la démonstration du théorème de représentation présentée dans Bolker [1967]. La première est une règle de moyenne (*averaging*) qui se compare à un axiome d'indépendance. Elle indique que si A et B deux options disjointes dans une structure de préférence, alors l'union de ces deux options est située à l'intérieur de l'intervalle qui les sépare. Autrement dit

(moyenne) $A \succ B$ implique que $A \succ (A \vee B) \succ B$ et de même

$A \approx B$ implique que $A \approx (A \vee B) \approx B$.

Il y a aussi une clause dite d'impartialité

(impartialité) Si A, B, C sont des options disjointes deux à deux, que

$A \approx B$ et $\neg (A \approx C)$ et que $(A \vee C) \approx (B \vee C)$ alors, pour tout D disjoint de A et B, $(A \vee D) \approx (B \vee D)$

Ces quatre conditions sur la structure de préférence suffisent pour établir le théorème d'existence qui assure l'existence d'une mesure de probabilité μ sur l'algèbre booléenne complète des propositions et d'une mesure signée ν sur la même algèbre telle que la relation $\succeq$ sur les préférences est représentée par la relation $\geq$ sur les quotients ν/μ de telle sorte que

$$A \succeq B \quad \text{si et seulement si} \quad \frac{\nu(A)}{\mu(A)} \geq \frac{\nu(B)}{\mu(B)}$$

Dans cette formule, les quotients de mesures ν/μ jouent le rôle de la désirabilité, c'est-à-dire l'utilité espérée au sens de Jeffrey[66].

Il y a cependant une difficulté qui se présente dans l'élaboration de cette théorie. Elle concerne l'application de la relation de préférence à des *propositions*. Considérons l'exemple qu'indique l'énoncé « Je préfère les poires aux pommes ». Ce serait faire violence à la grammaire que de dire que cette préférence est en réalité une préférence pour la proposition « Je mange des poires » comparativement à la proposition « Je mange des pommes ». Jeffrey nous dit que c'est Savage qui lui a donné la clef de l'énigme en proposant une interprétation simple et naturelle pour la notion de « préférer une proposition »[67]. Cette interprétation dit que préférer les poires aux

pommes, c'est accueillir plus favorablement la nouvelle (*to welcome the news that*) que je vais manger des poires. Ainsi, on pourrait paraphraser l'expression « l'attitude qui consiste à valoriser une option dans un problème de choix » par « l'attitude qui consiste à accueillir l'information selon laquelle le contenu propositionnel de l'option est réalisé ».

C'est un petit pas pour la sémantique, mais un grand pas pour la logique de la décision. En effet, cette interprétation permet de valider intuitivement l'unification sous un même type logique des objets des préférences et des probabilités.

> Si un agent délibère à propos de deux actes *A* et *B*, et qu'il est impossible d'accomplir *A* & *B*, il n'y a pas de différence entre la question de savoir s'il préfère *A* à *B* en tant que nouvelle ou en tant qu'acte, car c'est lui qui fait la nouvelle[68].

Cette unification a un prix cependant, car comme nous le verrons au chapitre suivant, le fait de poser une équivalence entre la valeur d'une option et la valeur *comme nouvelle* de cette option n'est pas sans conséquence. Elle permet un raisonnement à propos du problème de Newcomb que nous jugerons fautif[69]. L'autre volet de cette unification est la définition d'un acte comme proposition. Pour Jeffrey, un acte est « une proposition que l'agent a le pouvoir de rendre vraie »[70]. Il note que cette analyse d'un acte est grossière et qu'à proprement parler, un acte est plus qu'une proposition que l'agent tente de rendre vraie. L'importance de l'idée de tentative dans l'analyse de la délibération, importance que nous aurons l'occasion de souligner au chapitre VI, est reconnue par Jeffrey même si elle n'est pas intégrée à sa théorie.

4.3.3 Logique et structures bayésiennes

On peut présenter l'explication de la rationalité d'un choix chez Jeffrey comme une logique au sens propre, c'est-à-dire une théorie de l'inférence, en introduisant le concept de *structure bayésienne* (*bayesian frame*).

Définition : Une structure bayésienne est un quadruplet

$\langle \Omega, P, u, \mathbf{prop} \rangle$ où

Ω est un espace de mondes possibles,

P est une distribution des probabilités sur Ω,

u est une fonction qui assigne des valeurs d'utilité $u(\omega)$ aux différents mondes $\omega \in \Omega$,

prop est une fonction qui assigne aux constantes propositionnelles '*A*', '*B*',... des éléments de $\wp(\Omega)$. Ainsi, par exemple $A \subseteq \Omega$ est une proposition.

On définit le concept de *vérité* par la clause habituelle dans ce type de sémantique : A est vraie dans le monde ω si et seulement si $\omega \in A$.

Les connecteurs logiques &, $\vee$, $\neg$ sont représentés par les opérations ensemblistes $\cap, \cup, \sim$, respectivement l'intersection, l'union et le complément définis sur les sous-ensembles de Ω.

Dans cette logique, la *validité* d'une inférence est définie de la façon suivante : On dit qu'une inférence est valide si et seulement si la conclusion est vraie dans toutes les structures bayésiennes où toutes les prémisses sont vraies. Un énoncé est consistant si et seulement s'il existe au moins une structure bayésienne qui le vérifie.

L'intérêt des structures bayésiennes est double. D'une part, elles permettent de saisir le sens précis qui fait de la théorie de la décision une

logique de la décision. En ce sens, Jeffrey développe une idée phare de Ramsey, que nous avons déjà commentée, qui veut que la théorie des probabilités soit une généralisation de la logique formelle[71]. D'autre part, on pourrait montrer que les structures bayésiennes, malgré leur haut degré d'abstraction, ouvrent la possibilité de valider *théoriquement* le modèle standard par des considérations de symétrie. Il s'agirait de montrer que les structures bayésiennes assimilent toutes et uniquement les circonstances qui sont essentiellement les mêmes[72]. Pour réaliser cette validation théorique, on devrait utiliser la neutralité du modèle de Jeffrey pour les partitions, c'est-à-dire, en quelque sorte l'invariance relativement au découpage du monde.

Une structure bayésienne particulière (ou interprétée) précise le sens des composantes et permet de préciser la valeur d'un choix de façon « substantielle » et c'est pourquoi ce genre de spécification n'appartient pas à la logique « formelle » de la décision. On retrouve cette idée que le statut de la logique de la décision est comparable à celui de la logique déductive[73].

4.3.4 La probabilité conditionnelle

Comme nous avons eu l'occasion de le remarquer antérieurement, un agent qui délibère conformément au modèle de la délibération implicite dans les axiomatisations de Ramsey ou de Savage est confronté à un ensemble déroutant de possibilités qui inclut des options qu'il n'aurait jamais l'intention d'effectuer ou qu'il ne croit tout simplement pas possibles. Le fait de définir la valeur d'un acte, sa désirabilité espérée en fonction d'une valeur de probabilité qui est conditionnelle à la réalisation de l'acte permet de réduire ce vaste ensemble. C'est ainsi qu'il faut comprendre l'introduction de la probabilité

conditionnelle dans la formule de Jeffrey. Ce dernier améliore le modèle standard dont il a hérité mais il admet que la logique de la décision qu'il propose ne résout pas entièrement ce problème. Il ne se fait pas d'illusion sur le fait que l'espace de décision de sa théorie est encore trop vaste.

> [...] En effet, le cadre théorique bayésien est trop vaste car il admet des fonctions de croyance qui ne seraient acceptables que pour un fou, et des assignations de valeurs [de préférence] qui ne seraient acceptables qu'aux yeux d'un monstre.[74]

Bien que plusieurs recherches récentes établissent parallèlement que les probabilités conditionnelles constituent un objet de recherche qui a une importance décisive pour la poursuite du programme bayésien, la théorie de la décision (évidentielle ou causale) et la logique de la délibération, nous n'allons qu'effleurer le sujet qui pourrait faire à lui seul l'objet d'une recherche approfondie[75].

Nous allons commencer par dissiper deux malentendus. D'abord, contrairement à ce que certains traitements de la probabilité conditionnelle semblent suggérer, la probabilité conditionnelle ne doit pas être confondue avec la révision de croyances ou le changement d'opinion. La probabilité conditionnelle n'a pas un lien essentiel avec ces processus. Il s'agit tout simplement de relativiser la probabilité à une proposition ou à un ensemble de propositions. La probabilité conditionnelle s'explique par la règle du quotient

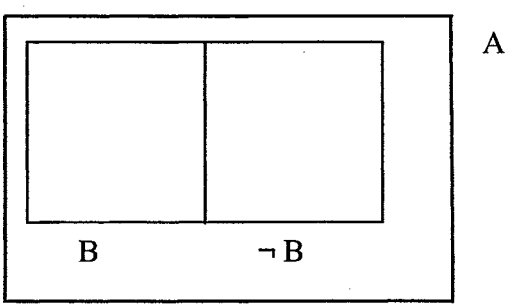
$$P(A \mid B) = \frac{P(A\,B)}{P(B)} \text{ pourvu que } P(B) > 0$$

L'autre malentendu que nous voulons dissiper concerne le statut de cette règle de calcul. Comme le note Jeffrey, ce n'est pas une définition. On peut s'en convaincre en observant que le ratio qui constitue le terme de droite de l'égalité est souvent indéfini. Par exemple, on pose que la probabilité que cette pièce de monnaie retombe sur pile si je la lance est de 1/2. Pourtant, le rapport de

$$\frac{P(pile \ \& \ pièce\ lancée)}{P(pièce\ lancée)} = \frac{l'indéfini}{l'indéfini}$$

Il faut donc considérer la règle du quotient comme un principe qui permet de relier la probabilité conditionnelle (à gauche) à la probabilité ordinaire (à droite). Comme l'indique cet exemple, il y a des circonstances où l'on utilise la notion de probabilité conditionnelle et où cet usage est intuitivement légitime même si la formule ne s'applique pas. Dit autrement, nous tenons à certains jugements de probabilité qui ne peuvent être le résultat de ce calcul.

Jeffrey nous donne une interprétation géométrique qui est éclairante. Si on représente l'ensemble $B \vee \neg B$ comme un rectangle, le calcul de $P(A \mid B)$ est la réduction du rectangle A à la portion B qui devient l'unité[76].



Un épisode fort intéressant de l'histoire de la logique contemporaine a conduit David Lewis à démontrer un résultat dit de trivialisation (*trivialization*) qui établit définitivement et sans appel que la barre « | » ne peut pas être un connecteur propositionnel, alors qu'on la prononce verbalement comme un « si » et qu'on l'interprète comme une sorte de conditionnelle. En plus de Lewis [1976] et [1986], on consultera Eells et Skyrms [1994] pour un aperçu de l'historique de cette question qui remonte au fameux test de Ramsey que nous discutions au chapitre II. Le résultat de Lewis établit que la barre « | » ne pourrait être un connecteur que dans un langage trivial, c'est-à-dire un langage dont la capacité expressive est ridiculement limitée, par exemple, ne comportant que des énoncés nécessaires ou impossibles[77]. Sans quoi, on peut montrer que la probabilité $P(A \mid B)$, somme toute, ne dépend pas de B. En effet, posons que X est une proposition contingente quelconque, alors, il suffit de poser que l'égalité $(A \mid B) \mid X = A \mid (X \ \& \ B)$ est vérifiée — et il est nécessaire qu'elle soit vérifiée si on veut que la barre « | » soit véritablement un conditionnel — et on peut montrer que la probabilité de A ne dépend pas du tout de la probabilité de B. La mise en abîme consiste en une démonstration que les propositions A et B sont indépendantes au sens des probabilités[78]. Comme l'observe Jeffrey, $P(A \mid B)$ ne serait alors qu'une façon maladroite d'écrire $P(A)$. Tout se passe comme si le signe « | » était une variante orthographique de la virgule. D'où la navrante conclusion : la barre « | » n'est pas un connecteur conditionnel. Vingt-cinq ans plus tard, les conséquences de cette conclusion se propagent encore comme une onde de choc dans tous les recoins de la logique des probabilités, de la logique de la décision et de l'épistémologie bayésienne. Peter Gärdenfors, par exemple, a proposé un analogue qualitatif du résultat de trivialisation de Lewis dans le contexte de sa

théorie de révision des croyances qui jette un doute sur la validité du *test de Ramsey*[79]. On se rappelle que le *test de Ramsey* que nous avons discuté brièvement au chapitre II correspond à la règle suivante si on l'expose dans le langage des conditionnelles de la logique épistémique :

> *test de Ramsey* : Acceptez un énoncé de la forme « Si A, alors C » dans un état de croyance K si et seulement si la révision minimale de K qui est requise pour accepter A exige aussi que vous acceptiez C.

La problématique de la révision des croyances est importante pour elle-même, mais elle n'occupe pas une si grande place dans la délibération orientée vers le choix rationnel.

À la suite de Gärdenfors, J. Joyce, de même que François Lepage ont proposé de contourner la difficulté par une sémantique ingénieuse pour le conditionnel subjonctif que l'on doit à D. Lewis, *l'imaging*. Joyce a étudié les relations entre *l'imaging* et la « conditionalisation ». On est tenté de conclure de ces recherches que l'irréductibilité de la probabilité conditionnelle est un fait premier, une donnée incontournable et irréductible de la logique du probable[80]. On peut penser, comme Hàjek que c'est la probabilité comparative qui est le concept plus fondamental et que tout jugement de probabilité doit être relativisé à une classe de référence. Par ailleurs, on peut chercher à élaborer une logique de la probabilité comparative qui prend la barre de conditionnalisation comme base, sans chercher à la réduire à un conditionnel[81].

4.3.5 Remarques sur l'ontologie propositionnelle

Dans la théorie de Jeffrey, les préférences, la désirabilité et les probabilités sont toutes définies sur le même ensemble d'options qui est constitué de propositions. Dans la terminologie de Jeffrey, un choix est une décision de rendre une proposition vraie. Une option dans une alternative est une proposition que l'agent peut rendre vraie, s'il désire qu'elle le soit[82]. Si on veut comprendre ce qu'est un choix, il faut donc clarifier ce qu'on doit entendre par proposition. Nous allons commenter la construction de l'algèbre L des propositions de la théorie de Bolker-Jeffrey. On considère l'ensemble des propositions fermé sous les opérations de disjonction et de négation (l'union et la complémentation). La définition de la conjonction est immédiate puisque $A \& B =_{déf} \neg (\neg A \vee \neg B)$.

Ainsi, L est une algèbre de Boole[83]. Elle contient deux éléments distingués, *T* et *F*. L'élément *T* est la proposition nécessaire (l'unité) et *F* est la proposition impossible (le zéro). Un *atome* est une proposition autre que *F* qui est impliquée par elle-même et par *F* mais par aucune autre proposition. Une des caractéristiques de cette algèbre est qu'elle ne contient pas d'atomes (*atomless*), autrement dit, son seul atome est *F* mais on retranche habituellement *F* de l'algèbre car la proposition impossible n'est jamais une option pour un agent dans un problème de décision[84]. On a en réalité $L - F$. Les éléments de l'algèbre sont toujours « moléculaires », chaque élément est un composé. Donc, aucun élément ne correspond à un seul monde possible. Comme le note Bolker, l'option « aller nager demain » peut se produire d'une infinité de façon, je peux aller nager seul, avec d'autres, la température de l'eau peut varier, et ainsi de suite[85]. Le postulat qui élimine les atomes

implique que toute option dans un problème de choix s'analyse comme une disjonction. Ainsi, par exemple, l'acte *A* deviendrait

$$(A \ \& \ C) \vee (A \ \& \ \neg C)$$

Contrairement à ce qui se passait dans la théorie de Savage, il n'y a pas de propositions pour jouer le rôle de conséquences, propositions qui posséderaient une désirabilité indépendante des désirs et des croyances, indépendante des jugements de probabilité. Ainsi, dans la théorie de Bolker-Jeffrey, dans les options constituées par des disjonctions, chaque terme propositionnel apporte sa valeur de probabilité, conditionnelle à la probabilité de la disjonction complète. Selon Jeffrey, c'est là un avantage de cette théorie comparativement aux autres,

> Je considère que c'est la vertu principale de cette théorie qu'elle ne fait aucun usage de la notion de pari ou de toute autre notion causale.[86]

L'ontologie des propositions qui est utilisée dans le théorème de Bolker a une cardinalité infinie dénombrable (*countable*). Or, Jeffrey ne cesse de répéter que les propositions de sa théorie peuvent être considérées comme des énoncés, à la manière préconisée par W.V.O. Quine. Au chapitre 12 de Jeffrey [1965] ainsi que dans Jeffrey [1974]b, Jeffrey indique clairement que les propositions s'interprètent comme des énoncés. L'usage d'*ersatz* linguistiques pour représenter les mondes possibles (comme récits maximalement consistants) a des vertus pédagogiques. Cependant, il y a une tension entre cette interprétation et l'algèbre de propositions qui est à la base de l'axiomatisation et de la preuve du théorème de représentation. On peut soutenir l'existence de cette tension par au moins deux arguments qui nous

semblent fort solides. Le premier argument est qu'on peut avoir des préférences pour des « état de choses », telles des saveurs, ou des nuances de couleur qu'on ne peut pas décrire en les formulant dans des énoncés; ces préférences peuvent être indescriptibles dans notre langage. Il y a plus de propositions que d'énoncés dans quelque langage que ce soit. En second lieu, on ne peut pas simplement décréter que l'algèbre des propositions est finie, car l'axiome de continuité est requis pour valider l'additivité dénombrable qui est utilisée dans la construction des mesures de probabilité ν et μ que nous discutions à la section 4.3.2. [87]. Jeffrey nous offre une interprétation qui reconnaît la difficulté tout en essayant d'en minimiser la portée. Dans Jeffrey [1965], en présentant ce qu'on doit comprendre comme un schéma d'application de la théorie, Jeffrey dit que l'on peut imaginer que la structure des préférences ne contient que les propositions qui sont exprimables dans le langage de l'agent[88].

> Un monde non-identifié, parmi tous les mondes possibles est le monde réel (*actual*) : il y a un récit (*novel*), parmi tous les récits cohérents et complets, qui raconte l'histoire vraie. En effet, nous connaissons même certains éléments de ce récit : ce récit appartient à chaque proposition vraie et on en connaît quelques-unes. (Le monde réel est identifié à l'ensemble de tous les énoncés vrais dans le langage de l'agent et nous savons que certains de ces énoncés sont vrais.) Mais tout ce que nous savons est compatible avec une infinité de mondes qui sont autant de candidats au statut de réalité.[89]

Les récits complets et consistants sont des représentations imagées d'une algèbre de Lindenbaum définie sur le langage de l'agent[90]. On peut ainsi passer de cette algèbre des propositions à une algèbre d'ensembles en vertu du

théorème de représentation de Stone qui établit l'existence d'un isomorphisme entre cette algèbre des propositions et une algèbre d'ensemble. C'est sur cette dernière qu'on veut définir les fonctions de probabilité admissibles[91]. Comme nous l'avons expliqué plus haut, l'ontologie propositionnelle et la théorie de la vérité qui est esquissée dans ce passage est standard en logique philosophique. Une des contributions de Jeffrey est d'avoir montré comment formuler la logique de la décision sur la base de cette ontologie propositionnelle standard, et ainsi, d'avoir rapproché le schème conceptuel de la théorie de la décision, telle qu'elle existait avant lui, de celui de la logique. On peut cependant déplorer que dans la théorie de Jeffrey, comme dans toutes ses rivales, l'appareil logico-mathématique soit plus explicite et mieux compris que la dimension philosophique.

4.3.6 Délibération et cinématique de la décision

Comme nous avons déjà eu l'occasion de le dire, la délibération au sens propre comprend tout ce qui se passe dans la tête d'un agent avant qu'il fasse un choix et qui se rapporte à ce choix. Jeffrey discute très peu de la délibération au sens propre. Il souscrirait sans doute à la formule de Skyrms :

> [...] on arrive ainsi à *la conception moderne dominante de la délibération* : On délibère en calculant l'utilité espérée.[92]

Il n'y a pas de doute que la logique de la décision de Jeffrey tente de valider ce que nous appelons *l'argument principal*, argument selon lequel une fois que les préférences et les probabilités subjectives sont déterminées, tout ce qui

importe pour fixer la valeur d'une option est déterminé. En principe, la délibération est une boîte noire dont on peut ignorer le mécanisme interne pourvu que l'on puisse cueillir une structure de préférence et une fonction de probabilité à la sortie. Cette façon d'envisager la délibération en logique de la décision ne correspond pas tout à fait à ce qu'en dit Jeffrey. Notre but dans cette section est de fournir une image probabiliste de la délibération qui soit plus détaillée que l'esquive qui correspond à l'argument principal à partir de l'explication de la cinématique de la décision.

Au début de la délibération, la probabilité de l'option choisie, disons p, se situe entre 0 et 1, c'est-à-dire qu'elle est strictement supérieure à zéro car on ne délibère pas à propos de l'impossible et elle est strictement inférieure à 1 pour des raisons tout aussi incontournables. D'une part, l'agent ne peut délibérer à propos d'actes qu'il serait certain d'accomplir et d'autre part, le produit de sa délibération est une intention d'accomplir un acte et non pas l'accomplissement de cet acte. Il existe toujours une possibilité pour que l'agent ne réussisse pas à accomplir ce qu'il avait prévu de faire. À mesure que l'agent délibère et qu'il se fait une tête, la probabilité de p va se modifier jusqu'au moment où l'agent va se former une intention résolue de faire p et alors la probabilité de cette option va s'élever pour approcher de 1[93]. Cette variation diachronique dans le degré de croyance doit cependant être cohérente à chaque instant, c'est-à-dire que tout accroissement de la probabilité de p, par exemple, doit s'accompagner d'un ajustement synchronique des valeurs de probabilité qui correspondent aux alternatives qui forment non p. Les fonctions de probabilité se répartissent sur des états doxastiques complets, une partition. Mais la délibération est compatible avec une fluctuation des états doxastiques pourvu que l'évaluation d'une option

dans une alternative soit référée à un état stable et cohérent des croyances de l'agent au moment qui correspond à l'instant du choix. La cinématique de la décision est suffisamment large pour accueillir les actions incertaines, les tentatives, de même que les jugements de probabilité incertains qui trouvent leur place dans le modèle de Jeffrey. En particulier, l'observation d'une variable aléatoire ne donne pas nécessairement une probabilité déterminée, mais peut résulter en une nouvelle distribution de probabilité selon une règle de mise à jour des croyances.

Il peut sembler que ces considérations sur la cinématique des degrés de croyance qui entrent dans le modèle du choix contenu dans la logique de la décision demeure bien en deçà de ce qu'on pourrait appeler une explication de la délibération. Aucune place spéciale n'est accordée au mode d'acquisition des connaissances, à la fixation de buts, à l'intentionnalité de l'action et tous ses autres ingrédients qui entrent dans les modèles cognitifs habituels de l'esprit d'un agent rationnel. Pour une part, la réponse du bayésianisme à cette sous-détermination de l'esprit est certainement explicable par ce que nous avons appelé *l'argument principal*. Du ce point de vue, peu importe la façon dont on modélise la délibération au sens propre, le *design* de l'esprit. Ce *design* ne va pas modifier la valeur d'un choix dans une logique de la décision pourvu que celle-ci soit suffisamment générale et abstraite. Mais Jeffrey pourrait aussi offrir une réponse plus substantielle à notre requête pour que soit clarifiée la délibération dans le choix rationnel. En effet, Jeffrey a élaboré les détails d'une épistémologie personnelle qu'il appelle le probabilisme radical. Il s'agit d'une épistémologie parce qu'il s'agit véritablement d'une théorie de la connaissance et parce que cette recherche prend comme point de départ les problèmes traditionnels de l'empirisme contemporain à propos de

l'induction et des paradoxes de la logique de la confirmation. Car Jeffrey, comme nous l'avons signalé au début de ce chapitre, est associé à la tradition de Hempel et de Carnap dont il fut d'abord l'élève avant d'égaler ses maîtres. On porte à son crédit des contributions substantielles à la méthodologie des sciences empiriques[94]. Son épistémologie personnelle à laquelle il donne le nom de probabilisme radical s'articule sur les apories du positivisme logique. Elle pose comme point de départ de l'épistémologie, le jugement de probabilité de l'agent. Ce probabilisme est dit radical car il n'a pas à reposer sur une base de certitude comme le probabilisme dogmatique de C. I. Lewis dans Lewis [1947] ni sur une distribution de probabilité antérieure à toute observation qui serait validée par une conception logique des probabilités comme celle de Carnap. Il n'y a pas de *définition* de la probabilité inconditionnelle, ou de règle impersonnelle générale. Comme dit Jeffrey, si on adopte le probabilisme radical, il n'y a que des probabilités, de bout en bout. Jeffrey désigne sa conception des probabilités par l'appellation « probabilité subjective » à la suite de De Finetti, mais il affirme qu'elle serait mieux décrite par l'expression *judgmental probability*. L'expression n'a pas d'équivalent français, mais elle exprime l'idée de jugement probable en évitant la connotation d'opinion (*doxa*) incertaine. La probabilité subjective que vous associez à un événement, tout en étant subjective, est néanmoins la probabilité que vous devez associer à cet événement, en tenant compte des données disponibles; ce n'est pas un simple caprice même si vous reconnaissez que quelqu'un d'autre pourrait associer une probabilité différente au même événement[95].

4.3.7 Jeffrey, Newcomb et le dilemme du prisonnier

Nous verrons au chapitre suivant comment la théorie de Jeffrey est vulnérable au fameux paradoxe de Newcomb. Jeffrey mentionne le paradoxe dans la seconde édition de Jeffrey [1965] et par la suite, présente et adopte pour un temps la solution dite « ratificationniste » pour contourner la difficulté. Il s'agit en quelque sorte de faire des choix qui pourraient être ratifiés par l'agent en fonction de ce qu'il sera après avoir choisi. La décision de faire un acte fait écran (*screens off*) à toute corrélation évidentielle qui pourrait exister entre les actes et les états du monde. Au cours de la délibération, si vous devenez convaincu que vous tenterez d'accomplir un acte, ceci peut devenir un argument pour changer d'idée et réviser votre intention initiale. La règle « choisissez les actes dont la désirabilité espérée est maximale » est ainsi remplacée par une règle qui est équivalente dans les cas ordinaires mais qui donne un meilleur conseil dans certains cas où la décision de choisir une option plutôt qu'une autre est perçue comme le signe qui révèle l'existence d'états de chose qui peuvent être désirables ou indésirables pour l'agent. Ce type d'approche découle de la recherche doctorale de Eells (1980) qui fut publiée sous Eells [1982]. Dans la seconde édition de *The Logic of Decision*, en 1983, il désavouera cette manœuvre qui ne résout pas tous les problèmes de décision de la famille du problème de Newcomb, ni le dilemme du prisonnier[96]. Examinons le dilemme du prisonnier. Le schéma de ce problème de décision est représenté par une figure à la page suivante. On vous a arrêté pour un crime que vous auriez commis et on présume que c'est avec un autre suspect qui est prisonnier lui aussi.

Les policiers qui ont construit cette énigme peuvent être assez confiants que vous allez avouer le crime et que l'autre prisonnier avouera aussi. Votre but est de passer le moins de temps possible en prison. Vous ne vous souciez pas du sort de l'autre et les seules conséquences prévisibles sont celles qui sont indiquées dans le tableau. De plus, vous croyez que les policiers vous ont dit la vérité quant aux peines que vous encourez. L'acte d'avouer le crime est l'acte dominant car sa désirabilité est supérieure peu importe ce que fait l'autre[97]. Vous devez considérer, et c'est ce qui construit le dilemme, que l'autre prisonnier est exactement dans la même situation que vous. Il peut aussi bien trouver que l'acte d'avouer est dominant et la conséquence dramatique est que vous serez tous deux pénalisés de quatre ans de prison si vous choisissez de façon rationnelle.

	L'autre n'avoue pas	L'autre avoue
Je choisis de me taire	1 an de prison	10 années de prison
J'avoue	Je suis libéré	4 années de prison

Le dilemme du prisonnier

Ainsi, l'acte qui semblait dominant ne semble plus aussi avantageux. Il serait plus avantageux de me taire si l'autre fait de même. Mais ce raisonnement,

l'autre aussi peut le faire. L'autre aussi veut minimiser ses années de prison et il ne se préoccupe pas de mon sort. La question devient : « Quel sorte de raisonneur est-il? ». On a dit que Jeffrey était sceptique à l'endroit du problème de Newcomb et l'on aura l'occasion de mieux comprendre pourquoi au chapitre suivant. Mais était-il également sceptique à l'endroit du dilemme du prisonnier? La réponse est oui, si on considère comme une position sceptique le fait de douter que les circonstances du dilemme du prisonnier constituent un réel problème pour la rationalité de l'agent. « Les dilemmes du prisonnier sont des circonstances cruelles, pas des pathologies de l'agent »[98]. La dynamique de la délibération est mise en évidence dans le dilemme du prisonnier parce que la valeur des options va varier selon la probabilité de l'acte choisi. Ma décision d'avouer devient une indication que cette option est également attrayante pour l'autre, donc moins attrayante pour moi. Nous verrons au prochain chapitre qu'il y a une circularité dans cette dynamique de délibération. Mais revenons à la position de Jeffrey. Le ratificationnisme recommande de faire un acte *A* qui est choisi parce que l'agent estime qu'il possède une désirabilité égale ou supérieure à toute autre option, en supposant que cet acte *A* sera effectivement accompli. B. Van Fraassen a imaginé un contre-exemple en modifiant l'histoire du dilemme du prisonnier de façon à ce que le meilleur acte ne soit pas ratifiable[99]. Dans cette variante, l'agent anticipe que certains facteurs pourraient l'empêcher de poser l'acte choisi, et pour cette raison, il se met à douter de l'action qui sera choisie par l'autre, celui-ci étant soumis aux mêmes facteurs. Ainsi, de part et d'autre, l'accomplissement des meilleurs actes échoue de façon corrélée (*correlated failures of execution*). Cette difficulté concerne vraiment la dynamique de la délibération. En effet, tout décalage entre le moment où l'option est évaluée et

le moment où se forme l'intention d'agir laisse place à la possibilité de contingences qui sont problématiques pour une logique de la décision. On peut imaginer cette même difficulté pour le problème de Newcomb. C'est le syndrome de la main qui tremble (*the trembling hand*). En théorie des jeux, on parle d'un équilibre de la main qui tremble lorsque des agents (joueurs) agissent non seulement en fonction de leurs croyances ordinaires, mais en tenant compte de croyances perturbées qui tiennent compte de la possibilité de petites erreurs. On peut considérer, par exemple, la personne maladroite qui est souvent incapable de faire l'acte qu'il a choisi de faire. Dans ces cas, sous certaines conditions, l'accomplissement de l'acte devient un meilleur indicateur des états favorables (la présence du million dans un problème de Newcomb) que la simple information révélée par le choix de l'agent. On peut alors montrer que l'acte ratifiable (refuser le mille dollars dans un problème de Newcomb) n'est pas le meilleur acte[100]. Par conséquent comme le note Jeffrey, la considération des actes ratifiables n'est pas un guide entièrement fiable pour déterminer la valeur d'une option (*choiceworthiness*).

La position finale de Jeffrey à propos du problème de Newcomb — en réalité, il s'agit d'un retour à sa position de départ — se retrouve clairement posée à la dernière page de son dernier ouvrage, Jeffrey [2004]. Il parvient à cette ultime position tranchée après avoir exploré diverses avenues pour aménager une place acceptable aux liens de causalité dans sa logique de la décision. Ses efforts théoriques dans cette direction furent exposés dans Jeffrey [1988], Jeffrey [1993] et Jeffrey [1996]. Tout compte fait, le jugement qui tombe est sans appel : le problème de Newcomb n'est pas un problème de décision selon Jeffrey mais un problème qui concerne ce que l'agent serait

disposé à faire, indépendamment de la délibération (*willi-nilly*). Nous examinerons l'argumentation de cette approche *no-boxer* au chapitre suivant.

CHAPITRE V

LE PARADOXE DE NEWCOMB ET LES THEORIES CAUSALES

5.1 Le problème de Newcomb et son interprétation

Il existe plusieurs variantes du problème de Newcomb et il n'est pas facile de discerner les versions de ce problème qui forment une classe d'équivalence. Nous allons d'abord présenter la version de référence de ce fameux problème, telle que proposée à l'origine par Robert Nozick, pour ensuite discuter quelques variantes qui ont le mérite de nous approcher du noyau essentiel de la difficulté[1]. Ensuite, nous allons discuter les principaux enjeux d'interprétation qui soulèvent des questions à propos de la délibération.

Vous êtes en face de deux boîtes. La première est transparente et vous pouvez voir qu'elle contient mille dollars. La seconde boîte est opaque et vous ne pouvez voir son contenu. Vous pouvez choisir de prendre le contenu de la boîte opaque uniquement ou vous pouvez choisir de prendre le contenu des deux boîtes. La difficulté est causée par la seconde boîte, celle dont vous ne pouvez voir le contenu. Cette boîte contient un million de dollars ou elle ne contient rien selon le résultat de la prédiction de quelqu'un dont vous pensez que les prédictions sont très fiables. Appelons ce prédicteur[2] un oracle. En principe, vous êtes libres de vous représenter ce prédicteur comme une divinité, un superordinateur ou un extra-terrestre si cela peut vous aider à

interpréter le problème. Si l'oracle a prédit que vous alliez prendre les deux boîtes, il n'a rien mis dans la boîte opaque. Si l'oracle a prédit que vous alliez prendre uniquement le contenu de la seconde boîte, il a mis un million de dollars dans cette boîte. Quel choix ferez-vous?

Il faut ajouter aux données de ce problème qu'au moment où l'agent délibère l'oracle a fait sa prédiction et le million de dollars se trouve déjà dans la boîte opaque ou ne s'y trouve pas[3]. De plus, il n'y a pas de piège. Si vous choisissez de prendre le contenu des deux boîtes, vous aurez le contenu des deux boîtes. Enfin, vous avez toutes les raisons de croire, par exemple, sur la base d'observations passées, que l'oracle est vraiment un prédicteur fiable. Il n'est pas nécessaire de présupposer une métaphysique déterministe. Il est suffisant que votre confiance dans la prédiction de l'oracle soit très élevée, par exemple que sa prédiction est juste à 99%. En d'autres mots, vous avez toutes les raisons de croire et vous êtes presque certain que l'oracle aura prédit correctement votre choix. Il faut bien avouer que le côté magique d'un oracle qui fait des prédictions quasi certaines peut dérouter un esprit rationaliste et sembler inintelligible. On peut et l'on doit tenter d'alléger ce côté magique en considérant l'oracle comme un psychologue perspicace (et riche) qui a bien étudié la personnalité de l'agent et qui s'appuie sur cette connaissance de la conformation de son esprit pour faire des conjectures à propos de son comportement, conjectures ayant un degré de probabilité élevé[4]. Cette interprétation dispose définitivement du côté en apparence magique de l'affaire. Mais ce n'est pas l'essentiel. L'essentiel, ce sont les trois conditions que nous avons énumérées et qui font partie de l'énoncé du problème ; ce sont des prémisses que l'on ne peut pas mettre de côté sans dénaturer l'énigme.

Voyons maintenant quelles sont les options. Vous pouvez délibérer en vous disant que l'oracle aura prévu le choix que vous ferez. Ainsi, vous n'aurez le million de dollars que si vous ne réclamez qu'une seule boîte. Avec ce raisonnement, vous concluez qu'il vaut mieux prendre le contenu de la boîte opaque uniquement. Si vous croyez que le meilleur choix est de prendre le contenu de la boîte opaque uniquement, vous êtes un raisonneur d'un certain type, vous êtes un *one boxer*[5]. On peut soutenir la valeur de cette option en raisonnant à partir de la maxime de l'utilité espérée:

(MUE) *Un agent rationnel confronté à un éventail d'actions possibles devrait choisir celle qui maximise l'utilité espérée.*

Mais vous pouvez aussi faire votre choix sur la base d'un autre raisonnement. Vous raisonnez de la manière suivante. Quelle que soit la prédiction de l'oracle, le million de dollars se trouve déjà dans la boîte opaque ou il ne s'y trouve pas. Ce que vous ferez maintenant ne peut pas influencer le contenu de la boîte opaque. Ou bien le million de dollars s'y trouve, ou bien il ne s'y trouve pas. Ainsi, en relation au moment présent et au futur, l'option qui consiste à prendre les deux boîtes est la meilleure. Si *x$* est le contenu de la boîte opaque, $x\$ + 1000\$ \geq x\$$ étant une proposition nécessairement vraie, l'option de prendre le contenu des deux boîtes est un meilleur choix. Si tel est votre raisonnement et que tout bien réfléchi vous croyez qu'il vaut mieux prendre le contenu des deux boîtes vous êtes un *two-boxer*. On peut défendre ce choix en s'appuyant sur le principe de dominance que nous avons déjà rencontré en exposant la théorie de Savage. On dit qu'une action *A* domine faiblement une action *B* pour un agent, si et seulement si pour chaque état du

monde, ou bien cet agent préfère les conséquences de *A* aux conséquences de *B*, ou bien les conséquences de *A* et de *B* lui semblent également préférables et pour certains états du monde l'agent préfère les conséquences de *A* aux conséquences de *B*. Ainsi, on peut soutenir la valeur du choix d'un agent *two-boxer* à partir du principe de dominance :

(DOMINANCE) *Soit une partition des états du monde telle que, relativement à cette partition, l'action A domine faiblement l'action B ; alors il vaut mieux faire A que B.*

L'exposé que nous venons de faire de cette énigme respecte de très près la formulation originale du problème de Newcomb dans Nozick [1970]. Comme le remarque Nozick, les étudiants et les amis à qui l'on soumet le problème et qui n'ont pas recours à un modèle théorique pour en faire l'expertise ont des avis divergents quant à la meilleure option dans ce problème. Cependant, parmi les philosophes et les théoriciens de la décision qui ont analysé cette version du problème de Newcomb, on ne trouvera vraisemblablement personne pour soutenir la viabilité de la solution qui consiste à ne prendre que le contenu de la boîte opaque (solution *one-boxer*)[6]. Autrement dit, le problème de Newcomb pose une difficulté pour une théorie de la décision comme celle de Jeffrey, une théorie qui recommande de maximiser l'utilité espérée de façon conditionnelle en supposant que l'acte est réalisé. Dans cette optique, l'évaluation des options suit la décision de réclamer, par exemple, le contenu des deux boîtes et le raisonnement du *one-boxer* s'applique inéluctablement : Si l'agent réclame le contenu des deux boîtes, l'oracle l'aura prévu et n'aura pas mis le million de dollars dans la boîte opaque. Dans

l'article où Nozick divulgue et propose la première discussion du problème de Newcomb, il écrit

> Étant donné la force des arguments en présence, il n'est pas suffisant d'arrêter son jugement en pensant que l'on sait ce qu'il faut faire. Il n'est pas suffisant non plus de répéter un des arguments, lentement et à voix haute. Il nous incombe de déconstruire l'argument opposé ; montrer pourquoi il ne tient pas tout en le prenant au sérieux.[7]

Convenons avec Nozick que si on en en reste au niveau des intuitions, il est difficile de surmonter l'impasse dans la confrontation des deux positions. De plus, il est clair que l'intérêt théorique du problème de Newcomb est de conduire à la confrontation de deux types de logique de la décision. Mais la situation devient inégale lorsque le problème est scruté à la loupe. Ce que nous aurons à dire à propos de la formulation du problème de Newcomb s'organise autour d'une idée directrice : Il ne faut pas conclure trop rapidement qu'une analyse minutieuse des données du problème ne nous apprendra rien. Au contraire, comme nous allons tenter de le montrer en explicitant chacun des raisonnements, on peut dégager certains présupposés qui impliquent une asymétrie dans la force respective des positions. De plus, nous allons chercher à montrer comment l'analyse du problème de Newcomb révèle l'importance de l'analyse de la délibération pour une définition du concept de choix rationnel. Ce faisant, notre objectif est de montrer que le problème de Newcomb révèle que le mécanisme de la délibération est un paramètre de la définition de la valeur d'un choix ; pour cette raison, la délibération doit être intégrée à la logique de la décision. On retrouve ici une des idées directrices de la présente recherche que nous avions mentionné au chapitre I. Cette idée

est sans doute plus facilement acceptable pour les partisans de théories causales de la décision que pour les tenants d'une théorie classique du type de celle de Jeffrey, théorie que l'on qualifie « d'*évidentielle* »[8].

L'analyse de l'argument *one-boxer* révèle quelques difficultés qui doivent être résolues pour que l'argument fonctionne. En premier lieu, il y a le problème des choix à effet rétrograde. Si je choisis de prendre les deux boîtes dit le *one-boxer*, l'oracle le saura et il ne mettra pas le million dans la boîte. Cette façon de formuler l'argument est incorrecte car quoi que je choisisse maintenant en vertu de l'autodétermination du contenu de ma volonté, le million de dollars se trouve ou ne se trouve pas dans la boîte; du côté de l'oracle, le coup est déjà joué. Ainsi, ma délibération juste au moment du choix ne peut pas altérer le contenu des boîtes[9]. Mon choix maintenant ne peut pas entraîner une action de l'oracle[10]. Par conséquent, la probabilité que j'assigne à la présence du million de dollars ne devrait pas être conditionnelle à la prévision de l'oracle[11]. Les gens qui n'ont pas exclu la possibilité d'une conditionnelle à effet rétrograde peuvent penser que le problème serait changé si quelqu'un était assis de l'autre côté de la table et qu'il pouvait voir le contenu de la boîte dite opaque, boîte qui dans cet arrangement ne serait opaque que sur les faces, qui sont dans mon champ de vision. Bien sûr, du point de vue de celui qui verrait le contenu des deux boîtes, peu importe ce qu'il voit, la solution *two-boxer* serait la bonne et le problème devient trivial. Mais du point de vue de l'agent qui doit choisir dans un problème de Newcomb, cette circonstance n'est pas censée faire de différence. Il n'y a pas de million de dollars qui apparaît ou disparaît de la boîte opaque à tout instant, selon le cours changeant de ma délibération. Ceci serait incompatible avec les données du problème tel que formulé. Le principe selon lequel le passé est

fixe et unique est un postulat fondamental de la métaphysique sous-jacente à la logique temporelle indéterministe[12]. Nous aurons l'occasion de souligner au chapitre VI qu'une logique de la décision qui ne serait pas compatible avec ce principe et d'autres principes de base d'une logique temporelle qui permet de situer les agents et leurs actions dans un monde indéterministe n'aurait pas beaucoup d'intérêt du point de vue philosophique. Pour rendre le problème intelligible, selon nous, on ne doit pas accepter la possibilité que la délibération au moment du choix ait un effet rétrograde sur les prédictions et les actions de l'oracle.

On trouve une discussion très détaillée des réactions spontanées au problème de Newcomb dans Nozick [1974]. Ces réactions sont d'autant plus intéressantes qu'elles proviennent de lecteurs du mensuel *Scientific American.* Le compte-rendu de Nozick fait état de positions réfléchies et informées utilisant parfois des arguments techniques ou présupposant une compréhension de quelques éléments de la physique contemporaine. En particulier, Nozick discute des relations entre la prédictibilité et la causalité inverse après avoir mentionné la position d'Asimov qui se dit déterministe[13]. L'énoncé selon lequel « la prédictibilité des décisions n'implique pas nécessairement le déterminisme » suggère deux possibilités. Ces deux possibilités correspondent à la causalité rétrograde (*backward causation*) et l'anticipation effective que permettrait une structure temporelle linéaire plutôt qu'arborescente. Nous ne voyons pas d'autre métaphysique temporelle qui permette l'anticipation effective, (*seeing ahead in time).* Malheureusement, ni la causalité rétrograde, ni la linéarité ne nous semblent acceptables ou cohérentes avec une métaphysique qui admet l'indéterminisme et la liberté[14]. Comme ces deux avenues sont bloquées, on n'a pas de modèle pour bloquer

l'implication. On peut donc soutenir que l'exactitude des prédictions de l'oracle, la prédictibilité du choix, apparaît comme un réel obstacle au libre-arbitre. Ceci ouvre la porte au rejet du problème en tant que problème de choix. Comment peut-on concilier la fiabilité des prédictions avec l'autonomie de la volonté dans la délibération ? Nous savons que la dimension métaphysique de la question est controversée, mais il nous semble que, tout bien réfléchi, il n'y a pas de choix s'il existe un fait à mon sujet qui *détermine* au moment de la prédiction de l'oracle que je vais, au moment du choix, choisir de prendre le contenu d'une seule boîte ou le contenu des deux boîtes. Pour contourner les débats métaphysiques, nous proposons un diagnostic prudent et conservateur. Dans un problème de Newcomb, l'agent possède une grande quantité d'informations sur le petit monde dans lequel il agit. À la limite, il sait tellement bien comment ses actions possibles sont corrélées à des états du monde qu'il peut prédire, qu'il devient difficile pour lui de se considérer comme un agent libre. Cette condition affecte la possibilité même de délibérer. Telle est selon nous, l'enseignement qu'il faut tirer des discussions nombreuses qui entourent la prédictibilité. Ces discussions sont peu fécondes. Il vaut mieux suivre Sobel et supposer que l'oracle fait des prédictions incertaines, comme un psychologue perspicace mais faillible.

Les observations qui précèdent permettent de comprendre comment on peut en arriver à soutenir, comme le fait Jeffrey à la fin de sa vie, que la position *no-boxer* s'impose puisque le problème de Newcomb n'est que le problème de savoir quelle sorte de personne je suis, indépendamment de toute activité de ma volonté libre et antérieurement à toute délibération. La difficulté semble avoir été anticipée correctement par Nozick lui-même ; il soulève la question : ne pourrait-on pas dire qu'une condition qui précède

universellement certaines décisions fait partie des effets de la décision plutôt que de la cause ? Encore une fois, si on rejette la causalité inverse, il semble qu'on détruit le problème de Newcomb en abolissant la possibilité d'un choix réel entre les deux options. Ceci, bien sûr, si par l'expression « choix réel », nous entendons l'exercice d'une volonté qui se détermine elle-même sans nécessité pour reprendre les termes kantiens.

On peut aborder autrement le problème posé par les prédictions. Pour rendre le problème de Newcomb intelligible, il faudrait donner un sens à l'idée que, sans influence causale, si ma volonté inclinait plutôt du côté de l'option de ne prendre qu'une boîte, l'oracle l'aura prévu et aura placé le million dans la boîte opaque. Puisque mon inclinaison ne cause rien d'autre qu'elle-même, un modèle plausible est que cette inclinaison est le signe, le présage ou l'auspice d'une condition préexistante chez moi que l'oracle aura su déceler et sur la base de laquelle il aura fait sa prédiction. C'est ce qu'on appelle un état profond ou une cause commune. Il existe une variante du problème de Newcomb qui met en évidence le phénomène de la cause commune ; c'est le problème de Fisher.

Le problème de Fisher est fondé sur une interprétation inhabituelle de la corrélation entre le cancer du poumon et le fait de fumer[15]. Dans l'explication de Fisher, le cancer du poumon est causé par une configuration génétique qui, par définition, est présente à la naissance. Les données du problème sont les suivantes : Le sujet qui a cette configuration génétique est plus susceptible de développer un cancer et il est plus susceptible de trouver que le fait de fumer est un choix qui domine celui de s'abstenir du tabac. De plus, le fait de fumer ne va pas augmenter la probabilité de développer le cancer, que le sujet possède ou non le mauvais gène. Dans cette interprétation, la consommation

de tabac et le cancer sont *indépendants* et sont le résultat d'une *cause commune* qui est la présence du mauvais gène. Le fumeur peut donc raisonner de la façon suivante : Si j'ai le mauvais gène, j'ai plus de chance de développer le cancer peu importe que je fume ou que je m'abstienne. Puisqu'il m'est plus agréable de fumer que de m'abstenir, alors le choix de fumer est un bon choix au sens de la règle qui veut que l'acte dominant maximise l'utilité espérée. Mais le second raisonnement que nous allons considérer est plus étrange car l'agent considère maintenant son envie de fumer comme une indication de la présence du gène qui cause le cancer : si j'ai envie de fumer, c'est le signe que j'ai le mauvais gène, donc en résistant à mon envie de fumer je diminue la probabilité d'avoir le mauvais gène ? Peut-on raisonner de cette façon ? Bien sûr que non, car le contenu de ma délibération actuelle, en supposant que je sois un agent libre de déterminer le contenu de ma volonté, ne peut pas modifier ma configuration génétique.

Un auspice est favorable ou fâcheux, selon la tendance préexistante qu'il révèle. Mais il est déraisonnable d'imaginer qu'on peut manipuler les effets pour dissiper les causes, ce qui serait comparable à la tentative d'éviter les hôpitaux pour ne pas être gravement malade, sous prétexte que le fait d'être à l'hôpital est un signe que l'on est gravement malade.

On évite ainsi le spectre de la logique de la décision vaudoue, pour reprendre une expression évocatrice due à Brian Skyrms. En essayant d'agir sur les auspices favorables, je ne devrais pas pouvoir produire des résultats qui vont dans le sens de mes attentes. Notons au passage que David Lewis a donné une caractérisation formelle de la classe de tous les problèmes de Newcomb, c'est-à-dire de problèmes où l'agent, abandonnant la proie pour l'ombre, adopte la politique de l'autruche. Cette caractérisation repose sur une

mesure du rapport entre la valeur *d'agir pour produire une nouvelle favorable* ou *défavorable* comparée au gain ou à la perte en elle-même[16]. Dans la prochaine section, nous verrons comment cette difficulté nous conduit directement aux théories causales de la décision. Les remarques de cette section avaient pour but d'établir que l'interprétation du problème de Newcomb nous oblige à une analyse minutieuse de l'énoncé du problème et soulève des questions d'interprétation qui sont difficiles.

Il est remarquable, par exemple, que le problème de Newcomb ne survive pas à l'idée de sa propre répétition, ce qui contraste avec le dilemme du prisonnier qui devient plus riche d'enseignements lorsqu'on considère la version répétée. La rationalité me dit de prendre les deux boîtes parce que cette option domine l'autre, mais je pourrais apprendre de mon erreur si j'étais déçu du résultat. Après tout, la disposition à apprendre de l'expérience est aussi constitutive de la rationalité[17]. Si, par exemple, j'habitais un pays où il y a beaucoup d'oracles qui proposent fréquemment à des gens que je connais des choix *à la Newcomb* et si je constatais que pour une raison qui m'échappe, les *one-boxer* réussissent mieux que les autres, je pourrais devenir *one-boxer* par pragmatisme. La rationalité continuerait de m'indiquer le choix *two-boxer*, mais le bon sens et l'expérience pourraient me permettre d'acquérir une disposition qui me permette de devenir riche[18]. Selon David Gauthier, il serait souhaitable que la « bonne règle », le choix rationnel, garantisse que la personne qui s'y conforme puisse s'attendre à maximiser son bonheur[19]. Mais ceci est un souhait. Dans une certaine mesure, ce type de raisonnement nous invite à sortir du cadre d'une théorie du choix rationnel fondée sur une conception bayésienne de la rationalité. Ce cadre théorique ne permet tout simplement pas de distinguer la rationalité d'un choix et son optimalité.

Nous poursuivons nos observations sur le problème de Newcomb en signalant un aspect moins connu de l'interprétation de ce problème, le phénomène de référence circulaire dans la délibération de l'agent.

Nous supposerons pour la suite de notre discussion que l'interprétation en termes de présage ou d'auspice n'élimine pas toute efficacité du processus de délibération, autrement dit, que le choix de l'agent est libre dans un univers indéterministe. Pour aborder le problème de circularité et écarter la difficulté associée aux conditionnelles à effet rétrograde, nous allons nous placer dans une perspective de simultanéité où non seulement l'agent délibère, mais l'oracle aussi. Dans cette version renforcée, il existe un phénomène de circularité dans la délibération qui est inéluctable. À notre connaissance, c'est à Brian Skyrms que revient le mérite de l'avoir signalé pour la première fois[20]. En présentant le problème de Newcomb, on pose habituellement que la prédiction de l'oracle est antérieure au choix de l'agent. C'est ce que semble indiquer la sémantique naïve du terme prédire : « dire avant ». Cependant, comme David Lewis l'a signalé, l'antériorité du moment de la prédiction n'est pas une caractéristique essentielle, pourvu qu'il y ait indépendance causale entre la prédiction et le choix de l'agent. Autrement dit, on peut interpréter la prédiction de l'oracle selon un autre sens du mot « prédire » qui est plus abstrait. On parle ainsi d'une théorie qui permet de prédire un événement, sans présupposer l'antériorité de la prédiction sur l'événement. Je pourrais dire, par exemple, que j'ai un modèle qui me permet de « prédire » le temps qu'il fera demain à partir de conditions climatiques d'hier même si la longueur du temps de calcul fait en sorte que la prédiction ne soit livrée que dans trois jours, deux jours après le jour comprenant les événements permettant de confirmer ou de réfuter ma prédiction. Lewis propose d'interpréter la prédiction de l'oracle

avec ce concept de prédiction. Ainsi, la prédiction de l'agent peut avoir lieu avant, simultanément ou même après le choix de l'agent. Cette observation permet à Lewis de rapprocher le problème de Newcomb du dilemme du prisonnier. Mais elle permet aussi de mettre en évidence un effet de référence croisée en formulant le problème de Newcomb d'une façon nouvelle. On peut poser que l'oracle simule le raisonnement de l'agent et que l'agent simule le raisonnement de l'oracle. La circularité est observée dans ce processus de délibération parallèle. Ce qui importe est que toute altération du raisonnement de l'agent, par exemple le fait de décider en cours de délibération de choisir de prendre plutôt une seule boîte que d'en prendre deux, va se répercuter sur ce que fera l'oracle, ce raisonnement pouvant être pris en compte par l'agent et ainsi de suite dans un jeu de miroirs sans fin. Les délibérations de l'agent et de l'oracle forment donc un processus de *simulation parallèle*. Peter Schorer a montré que ce mécanisme de *simulation parallèle*, ce qu'il appelle la « simulation parfaite », est en réalité une caractéristique commune à plusieurs paradoxes dont celui de l'examen surprise[21]. Il est clair qu'on retrouvera un cercle dans toute version du problème de Newcomb où l'acte choisi devient *eo ipso* une donnée du problème, du fait même d'être choisi. Comme l'observent Maitzen et Wilson, la question « Combien de boîtes prendrez-vous? » est en réalité une ellipse et la question complète est « Combien de boîtes prendrez-vous en sachant que l'oracle a prédit combien de boîtes vous prendrez? »[22]. Il y a clairement une régression sans fin dans cette délibération — en supposant qu'elle soit suffisamment minutieuse — car elle présuppose la connaissance de son résultat.

On serait tenté de conclure que l'effet de cette référence circulaire est de faire en sorte que l'une des clauses du problème est en réalité sous-déterminée

ou inintelligible. La conclusion vraisemblable serait que le problème de Newcomb n'est pas bien spécifié, qu'il est en quelque sorte un faux problème, parce qu'il camoufle un cercle vicieux. L'idée est attrayante et viendrait conforter une position *no-boxer*, comme celle de Jeffrey, pour qui, tout bien réfléchi, c'est le problème de Newcomb qui doit être rejeté et non la théorie bayésienne de la décision. Ce serait la leçon à tirer de l'énigme de Newcomb et l'on pourrait refermer le dossier. Mais il y a une conclusion différente que l'on peut tirer de cette difficulté. Le problème de circularité ouvre la perspective que le prédicat « ...est un choix rationnel » soit un prédicat problématique parce qu'il est logiquement *non-fondé*. Dans la délibération de l'agent, le phénomène de référence croisée fait que l'extension du prédicat varie selon les cycles de la boucle. Si l'oracle croit que je vais prendre le contenu d'une boîte seulement, alors le choix rationnel est de prendre le contenu des deux boîtes, mais alors l'oracle va ajuster sa prédiction et le choix rationnel deviendra celui qui consiste à ne prendre qu'une seule boîte, ce qui ferme la boucle. Clarifions le sens de l'expression « prédicat non-fondé » pour éviter les malentendus. La signification d'un prédicat classique peut être représentée par son extension ou par une fonction qui détermine pour chaque objet si ce prédicat est vrai ou non de cet objet. Un prédicat classique est fondé si son extension est définissable. On distingue parmi les prédicats dont la définition comporte une référence circulaire, ceux dont on peut définir l'extension. Eux aussi sont *fondés*. Le concept appartient à l'appareil logique requis pour traiter de la théorie des définitions en tenant compte des phénomènes de circularité[23]. La principale proposition démontrable qu'il faut avoir à l'esprit est que certains prédicats circulaires sont *fondés*. Dans certains cas, il est possible d'exprimer le sens du prédicat comme une règle de révision

qui permet de construire l'extension. Même si la définition du prédicat est cyclique, la règle de révision peut converger vers une extension stable. À la suite de Gupta, on peut considérer que le sens du prédicat est donné par une telle règle de révision et considérer que le concept de rationalité est un concept dont la définition générale est nécessairement circulaire, mais pas nécessairement circulaire de façon problématique. La question principale qui se pose est celle de savoir si le type de circularité qui existe pour le problème de Newcomb est de nature à ruiner le concept de rationalité lui-même. Cette orientation de recherche a été explorée pour certains problèmes discutés en théorie des jeux comme le problème du centipède et certains résultats ont été obtenus. On ne peut répondre avec assurance à la question principale pour la logique de la décision[24]. Si le concept de rationalité est défini sans restriction sur son domaine d'application, il est vraisemblable que ce concept soit essentiellement circulaire mais il est possible que cette circularité ne soit pas logiquement problématique. Ce diagnostic traduit précisément l'orientation de recherche annoncée de l'ouvrage collectif publié sous la direction de Chapuis et Gupta[25].

Une autre voie serait de chercher à énoncer explicitement les restrictions qui précisent la classe des problèmes pour laquelle le concept de rationalité est censé être légitime. Cette position nous semble raisonnable. Une explication du concept de rationalité qui restreindrait le domaine de son application pourrait bénéficier d'une explication de la délibération, plus précisément, d'un effort pour circonscrire le domaine du délibératif. En regard de nos propres buts, il faut accueillir le phénomène de circularité observé dans le problème de Newcomb comme une bonne nouvelle. Ceux qui croient que la rationalité est un concept circulaire, tout autant que les partisans de la position

no-boxer qui considèrent que le problème de Newcomb est inintelligible, pourront approuver la conclusion prudente que nous proposons. Selon nous, les difficultés posées par l'interprétation du problème de Newcomb ainsi que le phénomène de circularité que nous avons exposé invitent à clarifier le mécanisme de la délibération et s'inscrivent en faux contre toute perspective qui voudrait considérer l'énoncé traditionnel du problème de Newcomb comme transparent et non-problématique. Comme nous le verrons dans la deuxième partie de ce chapitre, la construction d'une théorie causale de la décision exige de rendre explicite certains aspects du raisonnement de l'agent qui concerne la façon dont il organise ses croyances à propos des options qui s'offrent à lui.

5.2 Newcomb et la théorie causale de la décision

Nous avons réservé un traitement séparé à l'analyse du problème de Newcomb selon la théorie causale de la décision parce que c'est ici que nous prenons congé de la théorie classique de la décision. Il fallait bien souligner ce passage où nous entrons dans un espace théorique où les thèses sont plus controversées et les débats plus récents. Comme nous l'avons déjà signalé au chapitre précédent, nous croyons que la position *no-boxer* de Jeffrey est cohérente, rationnellement justifiée, défendable. Les partisans de la théorie causale de la décision ont en commun de critiquer l'absence de considérations causales dans la définition de la valeur d'un choix que l'on trouve dans les théories évidentielles comme celle de Jeffrey. Nous verrons trois formulations différentes de la théorie causale. D'abord, celle d'Allan Gibbard et William Harper, les pères fondateurs de la théorie qui en ont donné la première

formulation en élaborant une suggestion de Stalnaker[26]. Ensuite, nous formulerons celle de David Lewis, qui est une version de référence dans la mesure où la plupart des autres versions s'énoncent comme des variantes de cette formulation[27]. Nous aurons l'occasion de formuler clairement la différence importante qui existe entre les théories causales et les théories évidentielles, en expliquant le rôle des partitions dans les théories causales. De plus nous établirons un contraste entre deux interprétations du principe de dominance qui met également en évidence la différence entre les deux types de théories. Enfin, nous présenterons les caractéristiques de la version de James Joyce qui est une version plus récente, plus robuste aussi, pour laquelle Joyce a donné un théorème de représentation. Les débats ne sont pas clos à propos de la question de savoir quelle est la meilleure formulation de la théorie causale de la décision. Cependant, nous croyons aussi qu'il existe, qu'il doit bien exister, une version de la théorie causale de la décision — version qui n'a peut-être pas encore été formulée de façon satisfaisante — qui serait suffisamment explicite et bien fondée. C'est l'ambition de Joyce, dans son ouvrage intitulé *Foundations of Causal Decision Theory*, de montrer que la théorie causale de la décision nous enseigne, tout compte fait, que la théorie bayésiennne de la décision de Jeffrey n'est qu'une partie propre d'une théorie plus générale qui serait la véritable logique de la décision.

5.3 Les théories causales de Gibbard et Harper et de Lewis

Voici la formule qui permet de calculer la valeur d'un acte, son utilité, dans la théorie causale de Gibbard et Harper [28]:

$$val(a) = \sum_i prob\ (\text{Je fais } a \boxdot\!\!\rightarrow r_i \text{ se produit}) \bullet (\text{dés } r_i)$$

La valeur de l'acte a, est la somme, pour chaque monde i, du produit de la probabilité que le résultat r_i se produise si j'accomplissais l'acte a par la désirabilité du résultat r_i. Gibbard et Harper ont représenté les clauses requises pour exprimer les dépendances causales à l'aide des conditionnelles dont l'antécédent est contraire aux faits, les contrefactuelles[29] : « Si j'accomplissais l'acte A, alors telle conséquence se produirait ». C'est cette conditionnelle que l'on exprime par le connecteur $\boxdot\!\!\rightarrow$. L'interprétation standard d'une conditionnelle de la forme $p \boxdot\!\!\rightarrow q$ est due à Stalnaker et Lewis. On présuppose qu'il existe un monde possible unique w_p qui est le plus proche du monde actuel et où p est vraie. La conditionnelle $p \boxdot\!\!\rightarrow q$ est vraie si q est vraie dans w_p. L'adaptation de cette sémantique aux conditionnelles de la forme $a \boxdot\!\!\rightarrow s$, où l'antécédent est un acte, est une simple restriction sur le type de l'antécédent. L'utilisation des conditionnelles contrefactuelles pour formuler une logique de la décision, en remplacement de la probabilité conditionnelle, apporte avec elle son cortège de difficultés. La sémantique en termes de mondes possibles de ces conditionnelles, telle qu'élaborée par Lewis dans Lewis [1973], exige d'ordonner les mondes possibles par une relation de proximité qui soit la moins arbitraire possible de façon à isoler le monde possible le plus proche ou la classe des mondes possibles équidistants les plus proches. Ceci représente une difficulté considérable et à ce jour, elle n'a pas été résolue de façon satisfaisante[30]. Cette difficulté est mise en lumière par des exemples comme la paire de conditionnelles suivante, entre lesquelles il est difficile de choisir :

Si Oswald n'avait pas tué Kennedy, quelqu'un d'autre l'aurait fait.

Si Oswald n'avait pas tué Kennedy, il aurait vécu centenaire.

Si on pouvait éviter les conditionnels contrefactuels et trouver un connecteur qui corresponde fidèlement à la définition de la probabilité conditionnelle,[31] on pourrait envisager la construction d'une théorie causale qui évite la sémantique des mondes possibles et reste plus proche du bayésianisme. La conditionnalisation représente la supposition du mode grammatical indicatif — on ajoute une proposition à notre base de connaissance — alors que la conditionnelle contrefactuelle représente le mode grammatical du subjonctif qui se conjugue avec l'imparfait. Malheureusement, comme l'a démontré Lewis en réfutant une suggestion de Stalnaker, il n'y a pas de connecteur raisonnable d'implication logique qui corresponde exactement à la conditionnalisation[32]. Enfin, la sémantique valide fatalement le principe du tiers exclus conditionnel $(P \mathrel{\square\!\!\rightarrow} Q) \vee (P \mathrel{\square\!\!\rightarrow} \neg Q)$ dont l'interprétation est problématique et la validité contestable[33]. Mais certains énoncés conditionnels subjonctifs ne sont ni vrais, ni faux, comme Lewis l'a bien montré[34]. Son exemple bien connu met en évidence une difficulté d'ordre logique

> Ce n'est pas le cas que si Bizet et Verdi étaient des compatriotes, Bizet serait italien; ce n'est pas le cas que si Bizet et Verdi étaient compatriotes, Bizet ne serait pas italien; néanmoins, si Bizet et Verdi étaient des compatriotes, ou bien Bizet serait italien, ou il ne le serait pas.[35]

Ceci correspond à la formule suivante qui peut paraître contradictoire où *C* et *I* représentent respectivement les propositions où Bizet et Verdi sont compatriotes et que Bizet est italien.

$$\neg (C \mathrel{\square\!\!\rightarrow} I) \,\&\, \neg (C \mathrel{\square\!\!\rightarrow} \neg I) \,\&\, (C \mathrel{\square\!\!\rightarrow} I \vee \neg I)$$

D'un point de vue sémantique, le problème est de savoir si le monde possible le plus proche du monde réel en est un où Bizet et Verdi sont italiens ou un monde possible où ils sont tous deux français. Il n'est pas possible de trancher d'une façon qui ne soit pas arbitraire. Il est tentant de conclure que ces conditionnelles, pas nécessairement toutes les conditionnelles contrefactuelles, n'ont pas de valeur de vérité, qu'elles sont neutres. Comme certains l'ont remarqué, il est clair qu'une sémantique probabiliste va valider le principe $P \mathrel{\square\!\!\rightarrow} (Q \vee \neg Q)$. En effet, en vertu de l'additivité sur les *P-mondes*, on a $Pr\,(Q) + Pr\,(\neg\, Q) = 1$. Ainsi, le conséquent est certain. Selon nous, ceci est plus une indication de la généralité de la sémantique probabiliste qu'une justification du tiers exclus conditionnel dans ce contexte. Nous croyons que la véritable conclusion qu'il faut tirer de cette difficulté dans le contexte de la logique de la décision — par opposition, par exemple, au contexte théorique de la sémantique des opérateurs modaux — est que les mondes possibles sont, par définition, trop détaillés pour être de bons candidats pour représenter les options dans un problème de décision.

Dans leur article, Gibbard et Harper ont formulé quelques lois gouvernant le connecteur $\square\!\!\rightarrow$, mais ils notent que son interprétation dans ce contexte, est loin d'être évident. Soulignons au passage que l'expression « ce contexte », désigne bien le contexte de la délibération, c'est-à-dire ce qui se passe dans la tête de l'agent avant qu'il ne fasse son choix. Les résultats que cet agent envisage pour les options qu'il évalue doivent s'interpréter dans ces mondes possibles qui se distinguent au plus du monde de l'agent par le fait que dans ces mondes, l'acte envisagé étant accompli des conséquences qui intéressent l'agent sont vérifiées ou ne le sont pas.

Axiome 1. $(a \,\&\, (a \,\Box\!\!\rightarrow s)) \supset s$

Axiome 2. $(a \,\Box\!\!\rightarrow \neg s) \equiv \neg(a \,\Box\!\!\rightarrow s)$

Conséquence 1. $a \supset [(a \,\Box\!\!\rightarrow s) \equiv s]$

Au départ, aucune restriction n'est posée sur les actes *admissibles* et les mondes *possibles* qui en résultent. Il est manifeste, par exemple, que pour éliminer les actes à effet rétrograde, (*branching toward the past*) il faut restreindre la classe des mondes possibles à ceux qui ont le même passé que celui dans lequel l'agent va accomplir l'acte car l'agent ne peut altérer le passé. Le second axiome implique qu'il existe un seul monde w_a qui correspond au résultat d'accomplir l'acte a au moment t. Si l'acte a est accompli, w_a est a. Avec cette clause, les lois élémentaires des conditionnelles ne s'interprètent pas aisément dans le contexte de la délibération[36]. L'atomisme logique qui est requis pour valider l'axiome 2 est commode mais il n'est pas naturel. Ne vaudrait-il pas mieux dire, à la suite de Prior[37], qu'un acte envisagé partage l'ensemble des mondes possibles en deux, ceux où l'acte est accompli et ceux où l'acte n'est pas accompli? Si l'acte est accompli, le monde réel où l'acte est accompli appartient au premier ensemble, sinon il appartient à l'autre. Intuitivement, par exemple, la délibération à propos de l'acte de caresser un chat s'évalue dans un ensemble de mondes possibles qui sont compatibles avec l'accomplissement de cet acte. C'est la proposition « le monde qui diffère au plus du monde actuel par le fait que l'acte y est accompli » qui est une description définie. L'indéterminisme et la prise en compte de la dimension temporelle ne s'accommodent pas bien de l'atomisme logique comme nous avons eu l'occasion de le remarquer au chapitre II en

discutant la théorie de Ramsey. Notons au passage que Sobel et Lewis ont proposé des versions de la théorie de la décision qui permettent d'éviter de considérer le monde qui résulte de l'action comme étant unique[38]. D'un point de vue logico-philosophique, nous croyons qu'il faut retenir la possibilité de définir l'utilité espérée sans s'appuyer sur une infrastructure logique qui commande d'analyser les options à partir d'une représentation de possibilités maximalement spécifiques[39].

Habituellement, lorsqu'on formule un problème de décision comme un tableau, la partition des actes est représentée par les rangs et la partition des circonstances est représentée par les colonnes. De façon plus abstraite, on peut considérer une partition comme un ensemble de propositions qui indexe les mondes. Avant de poursuivre la discussion des théories causales de Harper et Gibbard et de Lewis, nous devons poser quelques définitions indispensables.

La théorie de Lewis est fondée sur le concept de proposition. Intuitivement, c'est-à-dire en feignant d'ignorer les difficultés logiques que recèlent ces idiomes, on peut dire qu'on croit à des propositions et qu'on assigne des valeurs de préférence à des options, qui se définissent en termes de propositions. Chaque proposition est un ensemble de mondes possibles, ces mondes possibles où la proposition est vraie. C'est une fonction de croyance, notée C, qui assigne les valeurs de probabilité. Cette fonction est définie pour toute proposition X comme une somme :

$$C(X) =_{\text{déf.}} \sum_{W \in X} C(W)$$

Les valeurs de probabilité sont assignées à des mondes plutôt qu'à des énoncés ou à des événements. De plus, comme le souligne Lewis, la théorie qu'il développe n'a de sens que si on l'interprète dans une structure de modèle

où la cardinalité de l'ensemble des mondes possibles est finie, ce qui est une fiction qui simplifie la répartition des valeurs de probabilité[40]. La croyance conditionnelle :

$$C(X/Y) =_{\text{déf.}} C(XY)/C(Y)$$

où XY est l'intersection $X \cap Y$ des propositions X et Y, autrement dit la conjonction des deux. On dit que la fonction de croyance $C(X/Y)$ est obtenue de C en *conditionnalisant* sur Y[41]. C'est la définition de la probabilité conditionnelle que nous avons déjà rencontrée dans les chapitres précédents.

Une *partition* est un ensemble de propositions tel que une et une seule de ces propositions est vraie à chaque monde[42]. Il est parfois utile de considérer une partition des mondes dans lesquels une proposition donnée X est vraie. Une *partition de X* un ensemble de propositions tel que une et une seule de ces propositions est vraie à chaque X-monde. Si Z est une partition quelconque et X une proposition fixe, XZ est une partition de X. On utilise le concept de partition pour définir le concept d'option dans un problème de décision pour un agent. Ainsi, on aura une partition de propositions qui distingue chaque monde où l'agent agit de façon différente. En vertu de l'infrastructure logique proposée par Lewis, l'agent accomplit un *acte* en rendant vraie l'une ou l'autre de ces propositions, ce qui équivaut à réaliser une *option*. La seule contrainte est qu'il ne peut pas faire en sorte qu'une proposition soit vraie si elle implique mais n'est pas impliquée par une proposition de la partition. Suivant Lewis, nous réservons la variable A pour le domaine des options de l'agent.

On caractérise les connaissances de l'agent de façon simplifiée. Pour les besoins de notre infrastructure logique, on écarte toute possibilité d'ignorance relative en postulant pour lui des contraintes épistémiques fortes[43]. On suppose

que l'agent sait tout ce qu'il doit savoir à propos de la façon dont ce qui l'intéresse dépend causalement de ses actions. Il sait ce qui est sous son contrôle et ce qui est hors de son contrôle, et ce, en relation avec chacune de ses actions possibles. De plus, cette connaissance est certaine. Il sait également quels moyens prendre pour arriver à ses fins. Comme le note Lewis, une propriété remarquable d'un tel agent est que la délibération conditionnelle à l'accomplissement d'un acte ne lui apprendrait rien de nouveau. Pour un tel agent idéalisé, on appelle *hypothèse de dépendance* une proposition maximalement spécifique qui décrit comment tout ce qui lui importe sur la façon dont les états du monde dépendent ou ne dépendent pas de façon causale de ses actions. Les choses étant ce qu'elles sont, une seule de ces propositions doit être vraie et elle sont mutuellement exclusives ; donc elles forment une partition. Suivant Lewis, nous réservons la variable K pour le domaine des hypothèses de dépendance.

Dans le contexte des théories de la décision et en particulier des théories causales de la décision, on doit s'assurer que les partitions soient suffisamment fines pour représenter les choix, c'est-à-dire les actions et leurs résultats. On doit aussi s'inquiéter de ce que les partitions soient suffisamment exclusives pour former des alternatives et respecter l'additivité. En ce sens, les théories causales de la décision et les théories non-causales sont sensiblement différentes du point de vue conceptuel. Le concept fondamental de valeur d'un choix dans la théorie de Jeffrey, la désirabilité, se laisse définir d'une façon qui est invariante par rapport aux partitions qui définissent les options. Autrement dit, la valeur d'un choix n'est pas sensible à la façon de découper la totalité des événements qui constituent le monde. Or, la définition de la valeur d'un choix du point de vue de la théorie causale n'est pas indépendante

des partitions[44]. La théorie ne devient intelligible que si des contraintes sont précisées pour spécifier les conséquences des actes correctement, c'est-à-dire en reflétant les relations causales de façon appropriée. Une incontournable condition de cohérence stipule que les éléments des partitions qui définissent les options soient « causalement homogènes ». Ainsi, la représentation des dépendances causales pour une option ne peut comporter à la fois $a \mathrel{\Box\!\!\rightarrow} s$ et $a \mathrel{\Box\!\!\rightarrow} s'$ où s et s' sont incompatibles. La façon de spécifier les relations de dépendance causale et d'indépendance causale dans la représentation des options est différente selon les diverses versions de la théorie causale et c'est à David Lewis que revient le mérite d'avoir exposé de façon critique les diverses propositions qui ont été faites[45].

Comme le note Lewis, une relation causale ne se représente pas par une seule conditionnelle contraire aux faits mais par une liste de telles conditionnelles qui forme un schème de contrefactuelles (*a pattern of counterfactuals*)[46]. La taille d'un tel schème croit rapidement. Ainsi, si j'ai trois actions possibles et pour chaque action, trois états qui peuvent en résulter, un schème totalement détaillé comportera 3^3, 27 conditionnelles contrefactuelles. On appelle *conditionnelles causales* les conditionnelles contrefactuelles qui appartiennent à de tels schèmes. Pour interpréter la forme $a \mathrel{\Box\!\!\rightarrow} s$, on postule que les conséquences s des actes a appartiennent à une partition suffisamment riche pour pouvoir distinguer les diverses conséquences incompatibles des actes, chaque s étant spécifié de façon totalement indépendante de l'acte. Un *schème complet* est un ensemble de conditionnelles contrefactuelles indiquant les relations causales, une pour chaque option. Ces définitions permettent à Lewis de compléter sa reformulation de la théorie causale de Gibbard et Harper.

Dans leur théorie, les dépendances causales sont représentées par des schèmes complets. Ceci s'explique en partie par le fait que leur formulation tente de refléter la théorie de Savage où le contenu d'une option est indépendant de l'acte accompli. De plus, comme nous l'avons déjà observé, l'infrastructure logique d'une théorie de la décision doit assurer que chaque monde résultant d'un acte soit défini de façon suffisamment fine. Comme le note Lewis, la caractérisation des mondes résultant des actes comme une conjonction de schèmes complets semble exclure la possibilité de dépendance d'une autre nature que la dépendance causale. Lewis ne donne pas d'exemple de ce qu'il a en tête, mais il est assez aisé d'imaginer des exemples de dépendances non-causales. Ainsi, l'énoncé « Tous les organismes qui ont un cœur sont des organismes qui ont un rein » exprime une telle dépendance. Il y a une probabilité très élevée qu'un organisme qui possède un cœur possède un rein mais cette relation n'est ni causale, ni nécessaire. Elle n'est pas causale parce que ce n'est pas la présence du cœur qui cause la présence du rein. Elle n'est pas nécessaire parce qu'il est possible d'imaginer un *design* anatomique ou la fonction du rein n'est pas associée à un organe spécifique. De même, on pourrait imaginer deux mondes qui diffèrent au plus par le fait que dans le premier, le fait de pointer le doigt au ciel compte comme une prière mais pas dans le second. La relation symbolique qui fait qu'un acte ou une parole compte comme une prière est une relation de dépendance qui n'est pas causale. Ainsi, l'objection de Lewis concerne le fait que la théorie de Gibbard et Harper ne distinguerait pas les résultats des actes de la façon la plus fine qui soit concevable[47]. Il est difficile d'évaluer la force de cet argument en dehors de la résolution d'un problème de décision spécifique. Convenons que l'argument indique une difficulté potentielle.

La version que David Lewis propose de la théorie causale tient dans cette définition de l'utilité espérée U d'une option *A* (pour un agent à un moment) :

$$U(A) =_{\text{déf.}} \sum_K C(K)\ V(AK)$$

où $C(K)$ est la fonction de croyance (*credence*), appliquée à l'hypothèse de dépendance *K* et $V(AK)$ est la fonction de valeur (ou préférence) appliquée à la partition induite par les hypothèses de dépendance *K* sur les *A*-mondes. Comme ailleurs, le choix rationnel est l'option qui maximise l'utilité espérée. La formulation de Lewis est causale à double titre : d'abord parce que le contenu des hypothèses de dépendance exprime les relations causales et ensuite parce que les hypothèses de dépendance sont causalement indépendantes les unes des autres, mais pas nécessairement indépendantes du point de vue probabiliste. On voit que le trait distinctif de cette version de la théorie causale de la décision tient surtout à sa façon de définir les options par des hypothèses de dépendance. Il n'est pas étonnant que ce soit le point de comparaison principal que retient Lewis pour comparer et représenter les autres versions de la théorie. La théorie de Jeffrey n'a pas à être sélective quant au choix des partitions car contrairement à l'utilité espérée que nous venons de définir, la valeur de désirabilité calculée selon la théorie de Jeffrey ne varie pas d'une partition à l'autre. Par ailleurs, comme la fonction de croyance distribue les valeurs de probabilité sur des ensembles de mondes (les propositions), la théorie de Lewis peut se généraliser en utilisant un processus de révision de croyance connu sous le nom d'*imaging*[48]. L'imaging décrit comment redistribuer les probabilités en les concentrant sur un monde, avec une révision comportant le moins d'arbitraire possible. Cette règle de révision

est une alternative à la conditionnalisation avec laquelle on doit la comparer[49]. Il ne fait aucun doute que la révision des croyances à la lumière d'une nouvelle information est un mécanisme important de l'épistémologie personnelle. Elle constitue un aspect du processus cognitif de l'agent qui fait partie de la délibération au sens large.

Une autre façon de mieux saisir la différence entre la théorie évidentielle de Jeffrey et la théorie causale de la décision est de contraster leurs interprétations du principe de dominance. Comme on l'avait noté au chapitre IV, on peut dire que les partisans de théories causales de la décision adoptent et défendent le principe de dominance tandis que les partisans d'une théorie du type de celle de Jeffrey ont des doutes sur la légitimité de ce principe. Ainsi, on compare deux versions du principe de dominance, une pour chaque théorie. C'est cette explication, que l'on retrouve chez Sobel, Gibbard et Harper de même que chez Joyce, que nous présentons ici. Caractérisons d'abord la notion d'indépendance du point de vue évidentiel. Dans le contexte d'un problème de décision, il y a indépendance du point de vue évidentiel si et seulement si les propositions utilisées pour décrire les états associés aux options sont indépendantes des actes. Un agent ne considérerait pas que le fait d'accomplir ou de ne pas accomplir un acte constitue un fait probant (*evidence*) indiquant qu'un événement va se produire. Les croyances de l'agent, autrement dit, sa fonction de probabilité, ne serait pas modifiée si on lui disait comment il choisira de se comporter. Les définitions qui suivent distinguent l'indépendance évidentielle et l'indépendance causale; elles sont dues à Joyce[50].

Le principe de dominance avec indépendance évidentielle :

Soit $\{E_1, E_2, E_3,\ldots\}$ une partition d'événements que l'agent considère comme indépendants du point de vue évidentiel relativement à son choix entre les options A et B. Si l'agent préfère faiblement A à B, en symboles $A \succeq B$, étant donné E_j pour chaque événement E_j de la partition $\{E_1, E_2, E_3,\ldots\}$, alors il devrait préférer faiblement A à B.

On dit que deux événements sont indépendants du point de vue causal si et seulement si les propositions qui décrivent les états qui caractérisent les options sont indépendantes des actes. Mais on ajoute que les options sont indépendantes des actes si et seulement si les actes n'ont aucune efficacité causale sur les états du monde qui résultent de l'accomplissement de ces actes.

Le principe de dominance avec indépendance causale :

Soit $\{E_1, E_2, E_3,\ldots\}$ une partition d'événements que l'agent considère comme indépendants du point de vue causal relativement à son choix entre les options A et B. Si l'agent préfère faiblement A à B, en symboles $A \succeq B$ étant donné E_j pour chaque événement E_j de la partition $\{E_1, E_2, E_3,\ldots\}$, alors il devrait préférer faiblement A à B. De plus si l'une de ses préférences est stricte et que l'événement associé E_i est non-nul, alors il devrait préférer strictement A à B, en symboles $A \succ B$.

On constate que le premier principe valide la solution *one-boxer* alors que le second principe valide la solution *two-boxer*. Le second principe contient l'idée que les probabilités subjectives d'un agent dans un problème de décision possèdent une valeur déterminée de façon indépendante de son choix. C'est la clef de la solution *two-boxer*. Ces deux principes semblent fort

semblables, alors pourquoi préférer l'un à l'autre? Malgré la symétrie apparente, un *one-boxer* ferait valoir que les partitions en cause ne sont pas exactement comparables du point de vue conceptuel. L'indépendance causale hérite de l'obscurité inhérente à la notion de cause. L'indépendance probabiliste des théories évidentielles, de son côté, ne recèle aucune obscurité conceptuelle qui soit comparable. C'est pourquoi, comme nous avons eu l'occasion de le remarquer au chapitre précédent, Jeffrey considérait le fait d'éviter toute notion causale comme un mérite de sa propre théorie. Du côté de la théorie causale, on doit en principe relever le défi de rendre intelligible, de représenter et de clarifier la logique de la relation causale. On peut très bien construire une théorie causale de la décision qui explique comment les choix d'un agent ne sont rationnels que s'ils maximisent l'utilité espérée au sens de la théorie causale. Cependant la signification philosophique, ce qu'on pourrait appeler *la valeur explicative* de la théorie ne sera satisfaisante que lorsque sera fournie une analyse des relations causales que l'on retrouve dans un problème de décision et que ces relations seront expliquées de façon adéquate. L'analyse de la causalité présente de grandes difficultés théoriques même si les perspectives théoriques semblent ouvertes par suite de travaux comme ceux de N. Salmon, J. Pearl et C. Hitchcock qui ont renouvelé la problématique[51].

5.4 La théorie causale de James M. Joyce

La logique de la décision que présente James Joyce dans *The Foundations of Causal Decision Theory* est la version la plus récente de la théorie causale de la décision que nous discuterons. Elle se distingue par un théorème de représentation (le premier pour une théorie causale[52]) qui permet

de valider la théorie évidentielle et la théorie causale. Sur le plan philosophique, Joyce se distingue par sa critique du behaviorisme et du pragmatisme qui colorent les théories classiques de la décision, celle de Ramsey comme celle de Davidson et Suppes, par exemple. Ainsi pour Joyce, nos désirs ne sont pas réductibles à des *dispositions à agir*[53]. En proposant sa conception de la logique de la décision, Joyce refuse de réduire les préférences de l'agent aux préférences révélées par ses choix. C'est en cela qu'il s'oppose au béhaviorisme. Il ne croit pas non plus que la description de la structure des préférences d'un agent suffise à déterminer complètement ses désirs et ses préférences. Il n'y a pas de réduction théorique possible de la rationalité épistémique à la préférence rationnelle[54]. C'est en cela qu'il rejette le pragmatisme qu'il donne comme la « philosophie par défaut » des théoriciens de la décision. La non-unicité du théorème de représentation devient ainsi plus acceptable; en dehors du pragmatisme, cette caractéristique n'est plus un défaut.

Joyce ne croit pas que l'interprétation du choix en termes de paris tienne la route. En effet, la recherche de définitions opérationnelles n'a plus de sens en dehors du climat intellectuel du positivisme logique qui prévalait dans la décennie de 1920[55]. Elle ne permet pas, par exemple, de rendre compte de la rationalité de nos « raisons justificatives »[56]. En fin de compte, l'interprétation en termes de paris a principalement un intérêt d'ordre pédagogique, une sorte de validité à première vue et qu'elle ne résiste pas à une analyse plus minutieuse.

Joyce prend la structure des préférences (*preference ranking*) et non les probabilités comme base de sa théorie[57]. Comme nous avons déjà eu l'occasion de le remarquer, nous avons une intuition solide à l'effet que les

agents peuvent ordonner des états du monde qui résultent de leurs actions possibles alors que la possibilité d'estimer les probabilités de façon exacte dépasse les limites cognitives des agents. Pour cette raison, Joyce utilise des jugements de probabilité comparatifs pour construire son théorème de représentation[58]. De plus, si la structure de préférence d'un agent est incomplète, Joyce impose une exigence d'extensibilité cohérente. Cette solution que l'on trouve aussi chez Jeffrey, nous l'avions jugée raisonnable quand nous l'avions discutée au chapitre précédent. Dans l'approche de Joyce, on trouve donc un effort très marqué pour tenir compte, au moins en principe, de quelques intuitions solides à propos des capacités cognitives des agents. Comme nous l'avons souligné à plusieurs endroits dans les chapitres précédents, c'est une critique récurrente de l'usage de la théorie des probabilités qu'elle requiert, dans la plupart des versions de la théorie de la décision, que l'agent distingue des degrés d'incertitude avec une précision qui n'est pas réaliste. À notre avis, les positions philosophiques de Joyce sont légitimes et bien fondées. Elles suggèrent des orientations de recherche prometteuses. L'importance que nous accordons à la délibération nous rappelle que les croyances et les désirs qui influencent nos délibérations, les croyances et ses désirs qui influencent nos choix, sont antérieurs aux préférences révélées dans l'action. Il y a donc plusieurs raisons d'approuver les orientations de Joyce.

Les fondements de la logique de la décision, tels que déclinés par Joyce reposent sur l'idée que la théorie évidentielle de Jeffrey est une partie propre d'une logique de la décision plus générale, une logique de la décision de nature causale[59]. Cette inclusion peut s'interpréter de deux façons. En un sens,

tout partisan d'une théorie causale va reconnaître sans difficulté que la définition de la désirabilité au sens de Jeffrey indique le meilleur choix lorsqu'il n'y a pas de dépendances causales qui sont en jeu. En ce sens, la théorie évidentielle serait une logique de la décision qui est incomplète parce qu'elle ne s'applique pas aux problèmes du même type que l'énigme de Newcomb. C'était la position de Lewis, par exemple. Dans un second sens, et c'est le point de vue de Joyce, la théorie évidentielle, de Jeffrey, n'est pas du tout une logique de la décision, mais une « théorie de la valeur », théorie qui serait une composante nécessaire d'une logique de la décision. Comme il le montre, on peut asseoir cette théorie de la valeur sur les mêmes fondements que la théorie causale[60].

Examinons la critique que formule Joyce à l'endroit de la théorie ratificationiste. Choisir une action parce que son accomplissement donne une raison de penser que quelque chose d'agréable va se produire, voilà l'essentiel de la position évidentielle. Dans la variante ratificationiste, la délibération consiste à estimer la valeur d'une option relativement, non pas à l'estimation que je peux en faire maintenant, mais relativement à l'estimation que je ferais après avoir pris ma décision.

L'idée d'examiner la dynamique de la délibération dans un problème de Newcomb et d'en représenter les principes semble donc essentielle[61]. Comme nous le disions en discutant la position de Jeffrey sur le paradoxe de Newcomb à la fin du chapitre IV, c'est la réalisation de l'acte choisi qui, le cas échéant, me permettra de ratifier ma décision. Pour Joyce, le ratificationnisme était effectivement la meilleure réponse que la position évidentielle pouvait offrir au problème de Newcomb, la seule façon pour une telle théorie de valider le choix que tous considèrent comme le meilleur, celui de prendre le

contenu des deux boîtes. Pour Jeffrey et Eells, la capacité pour un agent d'anticiper ses propres choix fait écran (*screens off*) à l'influence de l'information que l'acte sera accompli[62]. Ainsi, cette information ne vient plus perturber l'estimation de la désirabilité des options. Si vous êtes perspicace nous dit Jeffrey, vous savez que votre délibération affecte vos croyances et vos désirs. Ainsi, vous devez choisir en fonction de la personne que vous prévoyez être après avoir choisi. Comme nous l'avons vu à la fin du chapitre IV, le contre-exemple de Van Fraassen est une mise en abîme pour le critère de ratifiabilité.

Joyce s'emploie à montrer que sur le plan formel, la théorie causale et la théorie évidentielle de la décision sont deux variantes d'une même équation générale. Examinons la façon dont il procède. En un sens, on se trouve ainsi à récapituler les différentes équations pour l'utilité espérée que nous avons rencontrées dans le chapitre précédent et dans ce chapitre. Le point de départ est une généralisation de la formule de Savage qui s'applique même lorsque la probabilité des états dépend des actes accomplis. La formule de Savage que nous avions vue au chapitre IV s'exprime de la façon suivante dans la notation de Joyce[63]

$$\textbf{SAVAGE:}\quad \mathbf{UE}(A) = \sum\nolimits_{S} \mathbf{P}(S) \cdot \mathbf{u}\,(\mathbf{r}[A, S])$$

où **UE** (*A*) est l'utilité espérée, ici inconditionnelle, *S* est une variable dont le domaine est constitué des états du monde et **u** (**r**[*A*, *S*]) est le résultat que *A* va produire si *S* est réalisé. Comme nous l'avons vu, pour Savage, la valeur de probabilité **P**(*S*) est rigide; elle ne varie pas en fonction des actes qui sont accomplis.

Dans un second temps, comme chez Jeffrey, on formule une équation, dite générale, où les actes A seront évalués en supposant qu'ils sont accomplis et les résultats seront pondérés par leur probabilité étant donné A.

$$\textbf{Équation Générale}: \quad \mathbf{UE}(A) = \sum_S \mathbf{P}(S \,||\, A) \cdot \mathbf{u}\,(\mathbf{r}[A, S])$$

UÉ (A) est l'utilité espérée de A. Contrairement à S et la fonction de probabilité $\mathbf{P}(S \,||\, A)$ représente le degré de croyance qu'un agent associe à la proposition S si A est accompli. L'équation générale se réduit à **SAVAGE** dans tous les problèmes où les états sont indépendants des actes. L'équation générale permet à Joyce de définir les théories évidentielles et causales par deux équations qui ne sont différentes l'une de l'autre que par l'interprétation donnée à $\mathbf{P}(S \,||\, A)$. On doit d'abord préciser le sens de $\mathbf{P}(\bullet \,/\, A)$. On utilise les deux barres obliques " / " et " \ " pour distinguer les deux interprétations. Le terme $\mathbf{P}(S \,/\, A)$ dénote la probabilité subjective de S conditionnalisé sur A, probabilité que nous avons définie par la formule habituelle

$$\mathbf{P}(S \,/\, A) = \frac{\mathbf{P}(S \,\&\, A)}{\mathbf{P}(A)}$$

On peut exprimer l'équation caractéristique de la théorie évidentielle de la décision, le calcul de l'utilité espérée (ou désirabilité) de l'acte A, par la formule suivante

$$\textbf{(Théorie Évidentielle)}: \quad \mathbf{UE}(A) = \sum_S \mathbf{P}(S \,/\, A) \cdot \mathbf{u}\,(\mathbf{r}[A, S])$$

Par contraste, on exprime la théorie causale de la décision comme une généralisation de **SAVAGE** et la fonction $\mathbf{P}(\bullet \setminus A)$ est une mesure des « tendances causales » selon l'estimation qu'en fait l'agent.

$$(\textbf{Théorie Causale}): \quad \mathbf{UE}(A) = \sum_S \mathbf{P}(S \setminus A) \cdot \mathbf{u}(\mathbf{r}[A, S])$$

L'utilité espérée devient une mesure de l'efficacité avec laquelle un acte *A* va entraînerait un effet jugé désirable ou indésirable. Ainsi, dans la théorie de Joyce **UE** (*A*) est appelée valeur d'efficacité de *A*. Il ne fait aucun doute que la contribution la plus importante de Joyce à la logique de la décision est son théorème de représentation et que le théorème de représentation de Joyce valide et contribue à rendre explicite la théorie causale de la décision. Cependant nous allons plutôt mettre en évidence une autre contribution de Joyce qui est fort importante et que nous n'avons pas vue discuter dans les commentaires publiés de son ouvrage[64]. C'est l'explication qu'il donne de la base commune à toutes les théories causales. Nous allons présenter et commenter cette explication pour conclure notre examen des théories causales.

Les différentes théories causales interprètent différemment la signification de la probabilité $\mathbf{P}(\bullet \setminus A)$, mais il y a un consensus à l'effet que cette probabilité doit représenter la façon par laquelle l'agent pense que ses actes vont *causer* des changements dans le monde. Dans la foulée de la tentative d'unification des théories causales entreprise pas Lewis, Joyce va révéler un ensemble de principes que toutes les variantes possibles de l'approche causale doivent avoir en commun. Comme nous le verrons, ces conditions ont des analogues dans la théorie évidentielle. Examinons le noyau

dur de ce concept de probabilité causale. Pour le caractériser, il faut d'abord définir quelques notions préliminaires.

Dans son ouvrage, Joyce utilise la même structure pour exprimer les théories de Savage et de Jeffrey. Cette structure contient l'ensemble ***C*** des résultats possibles (ou conséquences) des actes, l'ensemble ***S*** des états qui sont les résultats possibles et l'ensemble ***A*** des actes possibles. On part de ces ensembles pour former l'algèbre de base ***D*** = (Ω, ***C***, ***S***, ***A***), une algèbre de Boole qui est une σ-algèbre[65]. Ω est l'ensemble défini par les clauses

(1) les partitions ***C***, ***S***, et ***A*** ∈ Ω

(2) pour chaque $X \in \Omega, \neg X \in \Omega$

(3) pour tout $X_1, X_2, X_3, \ldots \in \Omega$, la disjonction dénombrable

$$\vee_j X_j = (X_1 \vee X_2 \vee X_3 \vee, \ldots) \in \Omega$$

(4) rien n'est dans Ω si ce n'est en vertu des clauses (1) à (3).

On peut appeler « problème de décision » une structure ***D*** = (Ω, ***C***, ***S***, ***A***), car tout ce qui importe pour bien le spécifier est contenu dans ***D***. Appelons *acte grossièrement spécifié*, une Ω proposition X telle que chaque élément de ***A*** implique X ou implique $\neg X$; X étant une disjonction d'éléments de ***A***. On dit que ces actes sont grossièrement spécifiés parce que l'ensemble des actes n'est pas muni de toute la structure qu'une théorie comme celle de Savage impose. Parmi les sous-algèbres que l'on peut définir à partir de ***D***, on distingue nommément Ω (***A***), la sous-algèbre des actes *grossièrement spécifiés* définie comme l'ensemble des Ω propositions qui peuvent s'exprimer comme des disjonctions d'éléments de ***A***.

P $(X \setminus Y)$ = P^Y (X) défini sur le produit cartésien $\Omega \times \Omega$(***A***) dont chaque couple est composé d'un état et d'un acte.

(1) Chaque $\mathbf{P}(\bullet \setminus Y)$ est une probabilité définie sur les propositions de Ω.

(2) $\mathbf{P}(X \setminus Y) = 1$ pour tout Y pour lequel $\mathbf{P}(\bullet \setminus Y)$ est définie.

(3) $\mathbf{P}(X \setminus Y) > \mathbf{P}(X \setminus Z)$ si et seulement si l'agent juge que Y va davantage favoriser causalement la réalisation de X que Z ne va favoriser causalement la réalisation de X.

On peut ainsi parler de causes qui favorisent et de causes qui inhibent. Un agent considère Y comme une *cause favorisant* X si et seulement si $\mathbf{P}(X \setminus Y)$ est plus grande que $\mathbf{P}(X \setminus \neg Y)$. Dans ce cas, on peut aussi dire que $\neg Y$ est une *cause qui inhibe* X. On définit ensuite les relations suivantes :

Y possède une *pertinence causale positive* pour X si $\mathbf{P}(X \setminus Y) > \mathbf{P}(X \setminus \neg Y)$;

Y a une *pertinence causale négative* pour X si et seulement si

$\mathbf{P}(X \setminus Y) < \mathbf{P}(X \setminus \neg Y)$;

Y a une *pertinence causale* pour X si et seulement si $\mathbf{P}(X \setminus Y) \neq \mathbf{P}(X \setminus \neg Y)$;

X est *causalement indépendant* de Y si et seulement si $\mathbf{P}(X \setminus Y) = \mathbf{P}(X \setminus \neg Y)$;

Y *cause plus efficacement* X que Z si et seulement si $\mathbf{P}(X \setminus Y) > \mathbf{P}(X \setminus Z)$.

Comme le remarque Joyce, si on remplace la barre " \ " par la barre " / ", on obtient des principes de probabilités conditionnelles qui sont valides du point de vue évidentiel. Ceci constitue un argument en faveur de la thèse qu'il veut démontrer, c'est-à-dire que la théorie évidentielle de Jeffrey ne se distingue formellement de la théorie causale que par cette opération de probabilité relative qui devient l'efficacité causale dans la formule caractéristique de la théorie causale. Ceci nous conduit à la proposition principale :

Proposition : Pour construire une théorie causale de la décision, il suffit de définir $\mathbf{P}(\bullet \setminus A)$ de manière à respecter les clauses (1) à (3) qui gouvernent $\mathbf{P}(X \setminus Y)$ et de valider les sept relations causales que nous venons de définir.

La proposition de Joyce n'est pas seulement utile pour rapprocher les théories évidentielles et les théories causales. Elle pourrait servir à montrer comment une logique de l'action qui comporterait un concept de cause —pour un sens du mot « cause » qui serait applicable à l'efficacité de nos actes — pourrait être jugée compatible ou cohérente avec une théorie causale de la décision[66].

La version de la théorie causale dont Joyce se réclame de celle de David Lewis et elle est comparable à celle-ci en plusieurs points. L'utilité espérée est calculée à partir de la valeur inconditionnelle des actes, comme chez Savage. Les dépendances causales sont représentées par des $\boldsymbol{K}$-partitions, comme chez Lewis et Skyrms, mais elles ne sont pas construites de la même façon. L'idée est cependant la même, chaque partition devant fournir toute l'information nécessaire pour déterminer précisément comment ce qui intéresse un agent dans un problème de décision dépend des actes qu'il peut poser. Nous n'examinerons pas davantage la théorie de Joyce et les difficultés reliées à sa manière particulière de définir l'imaging pour interpréter $\mathbf{P}(\bullet \setminus A)$ car de son propre aveu, son traitement de cette question n'est ni complet, ni satisfaisant[67]. Nous terminerons en commentant une caractéristique de la théorie de Joyce qui a de l'importance du point de vue de la délibération, la question de l'attribution de la probabilité des actes.

La théorie causale de Joyce partage avec la théorie évidentielle de Jeffrey l'idée que les états du monde, les actes et les résultats des actes peuvent être unifiés sous le type logique des propositions. De plus, Joyce a

relevé le défi de Savage et il a réussi à définir l'utilité espérée d'une façon qui n'exige pas de quantifier sur toutes les propositions du vaste monde, toutes celles qui se trouvent dans l'ensemble Ω de l'algèbre de base, mais seulement sur les propositions *W* qui sont pertinentes pour formuler les options dans un problème de décision[68]. Il décrit le processus de passage du « grand monde » au « petit monde » comme un processus de raffinement du cadre. Nous avions vu que Jeffrey jugeait cette caractéristique, l'homogénéité de son ontologie formelle, comme un progrès significatif, un mérite important de sa théorie par comparaison à la théorie de Savage qui distingue les actes des états. Dans son compte-rendu du livre de Joyce, I. Levi critique la théorie causale de la décision et en particulier la version de Joyce, car elle *exige* que les agents qui délibèrent fassent des prédictions à propos de leurs propres actions[69]. Imaginons qu'un agent se représente lui-même comme vraiment libre, autrement dit capable de vouloir ses propres actions dans un monde indéterministe; Levi soutient qu'un tel agent ne pourrait pas, de façon cohérente, tenter d'estimer les probabilités de ses actions futures. La délibération engloutit la prédiction (*deliberation crowds out prediction*)[70].

Dans Spohn [1977], Wolfgang Spohn exprime aussi l'idée qu'on ne doit pas accepter les prédictions à propos d'actions futures[71]. Selon lui, un énoncé comme « Il est improbable que je porte un pantalon court l'hiver prochain. » décrit en réalité une probabilité pour une situation de décision et sa paraphrase légitime est : « Je crois qu'il est improbable que je vais me retrouver dans une situation de décision où je trouverais que le mieux est de porter un pantalon court. » Selon Spohn, une logique de la décision peut éviter les difficultés de la théorie causale de la décision et fournir la bonne solution dans un problème de Newcomb pourvu qu'elle renonce à valider explicitement ou implicitement

la possibilité d'accorder des probabilités aux actions futures. À notre avis, un modèle plus explicite de la délibération permettrait de clarifier davantage ce débat. Nous observons aussi que l'attribution de valeurs de probabilité à des actes est une pratique courante, voire essentielle dans les constructions de la théorie des jeux et qu'on ne devrait y renoncer que pour des raisons très solides, plus fortes que celles avancées par Levi et Spohn.

Tout au long de ce chapitre qui nous a conduit au cœur de débats actuels sur les logiques de la décision les plus récentes, nous avons eu l'occasion d'apporter des arguments supplémentaires en renfort de notre idée directrice qui revendique l'importance de la délibération dans la détermination d'une décision rationnelle. Nous avons vu que le problème de Newcomb et le problème de Fisher ne pouvaient être analysés de manièere appropriée que dans le cadre d'une explication de la décision qui est sensible au mécanisme de la délibération, à ses contraintes et à ses modalités. Nous avons vu aussi que la théorie causale de la décision, avec son problème de représentation des relations de dépendances causales, exige que soient éclaircies des questions difficiles à propos de la nature des croyances de l'agent, du mécanisme de révision de ces croyances et de l'analyse des conditionnelles qui représentent les suppositions qu'un agent doit faire pour délibérer et choisir. Comme nous l'avons vu, en souscrivant à l'idée que le choix rationnel est celui qui maximise l'utilité espérée et en reconnaissant que l'utilité espérée se calcule par le produit de valeurs d'utilité et de valeurs de probabilité, on n'a pas encore complété l'explication du choix rationnel. La rationalité d'un choix est aussi déterminée par certaines modalités de la délibération et elle est sensible à la façon dont l'agent va découper le monde et y reconnaître des facteurs qu'il juge pertinents pour son choix.

CHAPITRE VI

LA DELIBERATION

6.1. Quelques concepts de délibération

Le dictionnaire *Webster* donne quelques sèmes essentiels du verbe « délibérer » dans la définition qu'il propose: soupeser dans l'esprit ; considérer les raisons pour et contre ; prendre en considération de façon réfléchie, évaluer. Pour sa part, le *Robert* décrit la délibération comme « un examen conscient et réfléchi avant de décider s'il faut accomplir ou non un acte conçu comme possible ». En cherchant à distinguer le choix de l'opinion, Aristote, dans *l'Éthique à Nicomaque*, a écrit quelques pages célèbres sur la délibération. Il en précise quelques particularités acceptées par tous : (1) La délibération concerne ce que l'on peut accomplir, ce qui est en notre pouvoir ; (2) la délibération concerne les moyens et non les fins ; (3) l'objet de la délibération et l'objet du choix sont identiques sauf que lorsqu'elle est choisie, la chose devient déterminée[1]. Comme le note McCall, la première clause doit être amendée quelque peu ; l'agent délibère à propos de *ce qu'il croit* qu'il est en son pouvoir d'accomplir. De même, la troisième observation doit recevoir une interprétation charitable pour être intuitive. Néanmoins, il y a là quelques intuitions incontournables ; la notion de délibération désigne l'activité mentale qui précède une décision. Comme on l'a vu dans les chapitres précédents, on se représente les options, on délibère sur leurs mérites et le choix se fixe dans

une décision. En philosophie contemporaine, il existe une conception assez précise de la délibération en philosophie de l'action et en philosophie de l'esprit. On délibère à propos de ce que l'on devrait faire ou s'engager à faire dans une situation donnée. Pour ce faire, chaque agent a certaines attitudes de base : des croyances, des désirs et des intentions premières. Lors d'une délibération, un agent — ou des agents, s'il s'agit d'une délibération collective — font des raisonnements, en particulier des inférences pratiques dont la conclusion est une décision (ou un choix) de faire (ou de ne pas faire) certaines choses. Nous disons que cette conclusion de type engageante et parfois directive possède une direction d'ajustement des choses à l'esprit. Tout ceci est assez précis mais il faut le comprendre comme une description de ce qui doit être expliqué par une théorie formelle.

Ultimement, les sciences cognitives et les neurosciences en diront peut-être davantage sur le processus exact, mais il devrait être clair pour tout philosophe que l'exercice mental qui conduit au choix ressemble d'assez près à ces intuitions lexicalisées ainsi qu'aux observations du Stagirite. Comme nous l'avons dit au chapitre I en expliquant le caractère normatif de la logique de la décision, la théorie de la décision doit se réconcilier avec nos intuitions à propos de la façon dont nous délibérons et faisons nos choix. David Lewis énonce ce souhait sous la forme d'un slogan optimiste où se mêle sans doute un peu d'ironie :

> Je crois que la psychologie du sens commun une fois systématisée, devrait ressembler passablement à la théorie bayésienne de la décision[2].

La conception de systèmes intelligents dans le domaine de l'intelligence artificielle a engendré un bon nombre de modèles de la délibération, au sens

large du terme. Mentionnons à titre d'exemple les recherches autour de l'architecture de planification STRIPS (*STanford Research Institute Problem Solver*) ou encore l'architecture OSCAR de John Pollock qui se propose d'élaborer une « théorie générale de la rationalité »[3]. Le mérite de ces architectures peut être évalué par la performance des systèmes dans lesquels ils sont réalisés ; il nous a semblé qu'on ne pouvait exporter ces *designs* d'esprit à l'extérieur de leur contexte de validation. Sortis de leur contexte de validation, ils ont parfois l'apparence d'une sorte peu recommandable de psychologie *a priori,* une psychologie de nature entièrement spéculative. Il devrait être clair que ces transpositions ne sont compatibles, ni avec la méthodologie des sciences empiriques, ni avec la méthodologie des sciences formelles[4]. En les étudiant, on est souvent frappé par l'impression qu'on nous raconte une histoire. Par contre, d'autres études poursuivies dans les départements d'informatique nous paraissent de grande valeur et sont très voisines des buts et des méthodes de la présente recherche[5]. Ainsi, Mehdi Dastani et ses collaborateurs de l'université d'Utrecht partent du postulat qu'il n'y a pas une façon unique de décrire le processus de délibération, ce processus pouvant être spécifié de diverses façons[6].

Comme nous l'avons vu au chapitre IV et V, les théoriciens de la décision ont tenu compte, dans une certaine mesure, de quelques aspects de la délibération dans leurs analyses. C'est le cas de Richard Jeffrey qui a écrit sur la cinématique de la décision, ainsi que de Brian Skyrms qui a écrit plusieurs articles et consacré un ouvrage à la dynamique du choix rationnel. Dans une perspective différente, Edward McLennen a également consacré un ouvrage original à la dynamique du choix rationnel[7]. Nous allons d'abord reconsidérer

la conception que Jeffrey se fait de la délibération, ou plutôt, de la cinématique.

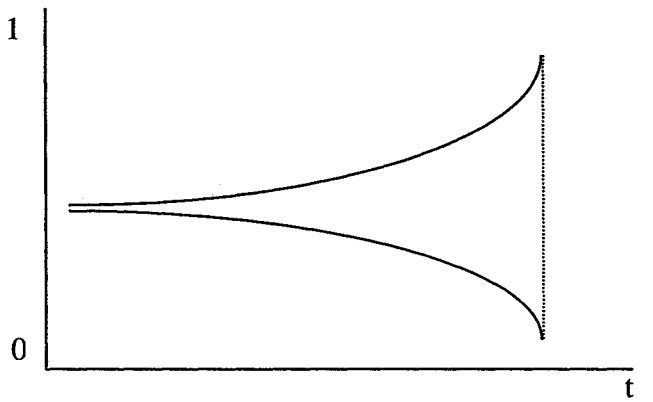


figure 6.1 la délibération : *Pr* évolue en fonction du temps

Au début de la délibération, la probabilité associée à une option se situe dans le voisinage de 0,5. Avec le passage du temps, la probabilité va augmenter pour s'approcher de 1 ou elle va diminuer pour s'approcher de 0. Elle n'atteindrait pas la valeur 1 parce que l'agent n'est jamais certain d'accomplir l'acte qu'il a choisi. Cette évolution est illustrée dans la figure 6.1 qui représente la fluctuation de la fonction de probabilité personnelle, par exemple, pour F (courbe ascendante) ou pour ¬ F (courbe descendante) en fonction du temps. Il va de soi que ces fonctions n'ont pas à être monotones bien que typiquement, elle le seront. La ligne pointillée indique le moment de la décision. Comme on le voit, Jeffrey propose d'exprimer dans un langage purement probabiliste une cinématique de la délibération. Il utilise le terme « cinématique » par analogie avec la discipline qui porte ce nom en physique ; la cinématique est plus abstraite par comparaison à la dynamique, elle décrit un monde plus idéalisé. Jeffrey et Skyrms ont aussi montré que la délibération

doit obéir à des contraintes. Nous allons voir la nature de ces contraintes à partir d'un exemple.

Considérons le problème de Fisher que nous avons exposé au chapitre V, où l'agent se demande s'il devrait fumer (F) ou non (¬ F). Le fait de fumer est une mauvaise nouvelle pour l'agent car il signale la présence du mauvais marqueur génétique, celui qui cause le cancer. La présence du marqueur est notée M et son absence ¬ M. De même, la présence d'un cancer est notée C et son absence, ¬ C. On peut exprimer une relation de dépendance en disant que ± M varie avec ± F et ± C. La cinématique de la délibération dans ce problème s'exprime par une série de contraintes sur la fonction de probabilité *Pr*. Dans la notation de Jeffrey, les valeurs des fonctions de probabilité s'expriment par des lettres minuscules. Ainsi, *Pr* (M)= *m*, *Pr* (C)= *c*, et ainsi de suite.

Un principe bien connu de la cinématique en physique est celui de la rigidité des solides et ce principe trouve son analogue dans la délibération. En effet, on trouve dans la cinématique de Jeffrey un principe de rigidité qui exprime la constance des probabilités conditionnelles durant la délibération. Notons au passage que cette stabilité des croyances à propos des relations de dépendance durant le processus de délibération est un sujet controversé dans la littérature récente[8]. L'agent délibère à propos de la question « Quelle serait la différence de désirabilité entre l'option de fumer et l'option de ne pas fumer si je découvrais que j'ai ou que je n'ai pas le marqueur génétique néfaste ? » Autrement dit, il veut estimer *des* (F) – *des* (¬ F) s'il découvre que M ou s'il découvre que ¬ M. Un agent qui délibère dans un problème de Fisher doit respecter six contraintes[9]. La première contrainte cinématique exprime l'idée d'un arrière-plan constant de probabilités conditionnelles :

Rigidité : Alors que $m = Pr$ (M) varie, les probabilités conditionnelles suivantes ne doivent pas varier :

$f = Pr$ (F | M), $f' = Pr$ (F | ¬ M), $c = Pr$ (C | M), $c' = Pr$ (C | ¬ M)

La seconde contrainte qui s'applique à la délibération dans le problème de Fisher exprime la corrélation entre le cancer et le fait de fumer :

Corrélation : si $p = Pr$ (C | F) et $p' = Pr$ (C | ¬ F) alors $p > p'$

Les deux contraintes qui précèdent, soit la rigidité et la corrélation expriment ensemble un concept probabiliste de dépendance causale[10]. La troisième contrainte découle de ce qu'on ne délibère pas de ce qui est certain ou impossible.

Indétermination : Aucune des valeurs f, c, f', c' n'est égale à 0 ou 1

Le principe suivant est familier

L'indépendance : Pr (F & C | M) = fc et Pr (F & C | ¬ M) = $f'c'$

À partir des contraintes d'indépendance et de rigidité, on peut montrer (en posant une égalité qui est presque toujours vérifiée[11]) que le marqueur fait écran entre le fait de fumer et le fait de développer le cancer. Il n'y a plus de corrélation directe entre F et C.

Faire Écran : Pr (F | C & M) = a , Pr (F | C & ¬ M) = a'

Pr (C | F & M) = b , Pr (C | F & ¬ M) = b'

Dans le problème de Fisher, contrairement au problème de Newcomb, la délibération n'affecte pas l'influence des « états profonds » sur les actes et les choix[12]. Même si on n'utilise pas le langage causal pour les décrire, il est clair que les conditions M et ¬ M sont utilisées comme des hypothèses causales. Le langage causal ne joue pas de rôle dans cette explication nous dit Jeffrey, ce n'est qu'un commentaire. Tout compte fait, le calcul de Jeffrey pour le problème de Fisher donne F comme l'acte dominant, comme acte pur ou

comme acte mixte en présence de M. C'était bien le sens de l'argument de Fisher[13]. Telle est la contribution de Jeffrey à l'analyse de la délibération. Nous avons deux remarques à faire à son sujet. La première concerne les limites de la délibération et la seconde concerne l'approximation probabiliste de la relation causale.

Il y a une idée-phare qui revient dans tous les textes où Jeffrey discute le problème de Fisher ou le problème de Newcomb. Ici, l'agent ne délibère pas tant à propos de ce qu'il devrait faire, mais de la sorte de personne qu'il est ; il s'inquiète de savoir s'il a un marqueur génétique néfaste, il s'inquiète de savoir s'il est, de façon prévisible par un oracle, le genre de personne qui prendrait le contenu des deux boîtes dans un problème de Newcomb. Cette façon de formuler les choses pourrait nous conduire à rejeter ces problèmes hors de la classe des problèmes de décision ; la logique de la décision ne s'appliquerait pas à ces problèmes. À notre avis, pour ce cas à tout le moins, ce serait une erreur car un grand nombre de problèmes de décision peuvent se paraphraser de cette façon. La plupart de nos décisions morales du genre « Il n'est pas bien d'abandonner un ami alors qu'il a besoin d'aide, donc je vais aider mon ami » peuvent se réécrire sous la forme : « Bien sûr, je suis occupé, mais *suis-je la sorte de personne qui* abandonnerait un ami alors qu'il a besoin d'aide ? »[14]. Il nous semble qu'il n'existe aucune raison de refuser la légitimité de cette paraphrase[15]. Nous croyons que la délibération d'un agent s'effectue sur l'arrière-plan de ses croyances et qu'il n'y a pas de raison pour exclure *a priori* ses croyances au sujet des états profonds (*deep states*).

Il y a aussi une difficulté avec la représentation probabiliste des relations de dépendances causales. Les relations de dépendances causales exprimées par des contraintes sur le processus de délibération sont des

dispositions de l'agent à l'endroit de l'évolution de ses propres croyances. Ceci est particulièrement clair dans le cas de la contrainte de rigidité. Elle n'est pas vraiment l'expression d'une croyance à propos d'une dépendance causale (une constance dans la dépendance entre les effets et les causes) mais bien plutôt une contrainte sur la stabilité des croyances à propos de l'arrière-plan. Jeffrey et Skyrms reconnaissent que la contrainte de rigidité ne s'applique pas dans un problème de Newcomb. À notre avis, ils reconnaissent ainsi que les dépendances causales ne sont pas bien représentées par les dites contraintes. Il est possible que l'on parvienne un jour à donner une description fine des relations causales dans un langage probabiliste[16]. Cependant, nous avons de solides raisons d'en douter, pour des raisons qui remontent aux arguments de Hume. Comme on l'a remarqué au chapitre IV, dans l'explication de la délibération de Jeffrey, on ne trouve que des relations logiques qui indiquent diverses façons par lesquelles la vérité d'une proposition tend à promouvoir la vérité d'une autre[17].

Notre suggestion est que la logique de la décision doit établir le langage causal à un autre étage, à l'étage de ce que nous avons appelé l'infrastructure logique[18]. La logique de la décision devrait se construire sur un langage qui a la capacité expressive pour situer les agents, leurs délibérations, leurs actions et les états qui en résultent. Un tel langage doit permettre de résoudre les difficultés au niveau où elles se posent. Nous croyons avoir montré au chapitre précédent qu'il est nécessaire d'intégrer une logique temporelle de l'action pour éliminer les conditionnelles à effet rétrograde. Nous croyons aussi avoir établi la difficulté de résoudre le problème du conditionnel de la délibération dans le cadre limité de l'algèbre des actes et des états. Dans la prochaine section, nous allons expliquer cette orientation de recherche en exposant dans

le cadre d'une logique temporelle indéterministe où l'on peut représenter la nécessité historique, l'indéterminisme, la décision et le choix[19].

6.2. La logique du temps ramifié et la logique de l'action

Aux chapitres II, IV et V de la présente thèse, nous avons eu l'occasion de déplorer le fait que l'infrastructure logique nécessaire pour rendre explicite les divers aspects de la délibération dans les théories de la décision était généralement restée dans l'ombre, sous-développée ou sous-déterminée par ces théories. Dans cette section, nous nous proposons d'essayer de remédier à cette situation et de décrire un cadre formel qui nous semble essentiel pour relier conceptuellement les divers aspects du choix rationnel. L'unification des logiques sous-jacentes à la logique de la décision ne sera pas une tâche facile et nous ne prétendons pas l'avoir réalisée. Nous devons cependant défendre la nécessité de cette unification et déplorer son inexistence de façon explicite. Dans le même esprit que McCall [1994] et de Horty [2001], nous croyons que la décision et la théorie normative du choix ont comme arrière-plan logique une logique temporelle indéterministe et une explication de l'action qui situe les agents et leurs actions dans cette logique. Historiquement, les idées fondamentales de la logique temporelle indéterministe sont dues à A. Prior et R. Thomason[20]. Plusieurs autres logiciens en ont inspiré le développement, tels L. Åqvist, et B. Chellas. Plus récemment, d'autres ont contribué à la rendre explicite et détaillée tels S. McCall de même que Nuel Belnap et ses collaborateurs[21].

La logique temporelle indéterministe a servi de contexte formel pour développer une logique de l'action, connue sous le nom de « logique du *stit* »

(*stit logic* pour *seeing to it that*), unifiée dans l'ouvrage de Belnap, Perloff et Xu *Facing the Future*[22]. Cette logique de l'action s'inscrit aussi dans une longue tradition d'efforts philosophiques pour expliquer la structure logique de l'action, tradition qui remonte jusqu'à St-Anselme. Un résumé de la position philosophique de cette conception des agents, de leur liberté et du monde indéterministe est donné dans Belnap [2002]. Daniel Vanderveken a aussi proposé une logique de l'action et de la temporalité qui présente des similitudes, mais qui s'appuie sur sa théorie prédicative des propositions et sur une conception intentionnelle de l'action qui est significativement différente. La sémantique de la théorie présentée dans Belnap, Perloff et Xu [2001] est la version la plus détaillée d'une logique de l'action pour des agents qui font des choix dans un monde indéterministe, nous la prenons comme référence[23]. Pour abréger l'expression « logique temporelle indéterministe » et pour désigner la famille des logiques qui s'inscrivent à la suite des théories de Prior et Thomason , on utilise l'expression « logique du temps ramifié » (*the logic of branching time*)[24]. Plusieurs théories différentes peuvent être caractérisées comme des variantes ou des extensions de ce modèle. Nous allons décrire sommairement la sémantique de cette logique temporelle qui est une partie importante de l'infrastructure logique d'une dynamique de la délibération. Notre but est de décliner un long préambule destiné à rendre intelligible nos remarques sur les rapports entre la logique de la décision et cette infrastructure logique. Nous présentons la sémantique telle qu'on la trouve dans l'article de Horty et Belnap [1995] et dans Horty [2001], un ouvrage qui développe l'aspect normatif du choix et qui propose l'infrastructure logique requise pour exprimer une version générale de l'utilitarisme des actes. Nous verrons ensuite comment situer la décision et le choix délibéré dans ce modèle[25].

En logique intensionnelle, il est habituel de définir la vérité d'une proposition relativement à un index. Dans la sémantique de la logique modale due à Kripke, cet index est l'ensemble des mondes possibles. Dans la logique temporelle proposée par Thomason, cet index était composé à partir des mondes possibles W et des moments du temps T ; c'est le modèle $W \times T$. Par contraste, la logique temporelle de Belnap rend la vérité de chaque proposition relative à un moment m et une histoire h à laquelle ce moment m appartient. À titre d'exemple, on dira que la proposition P est vraie à m_1 / h_2. En partant de postulats simples comme l'idée que le passé est déterminé et qu'il ne peut pas changer tandis que le futur est ouvert pour l'action et le hasard objectif — la chance — il est naturel de se représenter les moments du temps comme étant ordonnés dans une structure qui prend la forme d'un arbre, comme dans la figure 6.1. La structure de l'arbre indique déjà l'impossibilité de l'action sur le passé dont nous avons dit qu'elle était fort importante pour éliminer une interprétation indésirable du problème de Newcomb. Dans la représentation en arbre, un même moment peut appartenir à plusieurs histoires à la fois. On peut définir un grand nombre de concepts utiles dans ce contexte de la logique du temps ramifié. Nous nous limiterons à considérer ceux qui nous conduisent aux concepts fondamentaux de la délibération.

Le point de départ habituel est la définition d'un arbre : c'est un ensemble de moments qui sont ordonnés par une relation d'ordre ≤ partiel [26]:

réflexive : $(m \leq m)$

transitive :$(m_1 \leq m_2 \ \& \ m_2 \leq m_3) \supset m_1 \leq m_3$

antisymétrique : $(m_1 \leq m_2 \ \& \ m_2 \leq m_1) \supset m_1 = m_2$

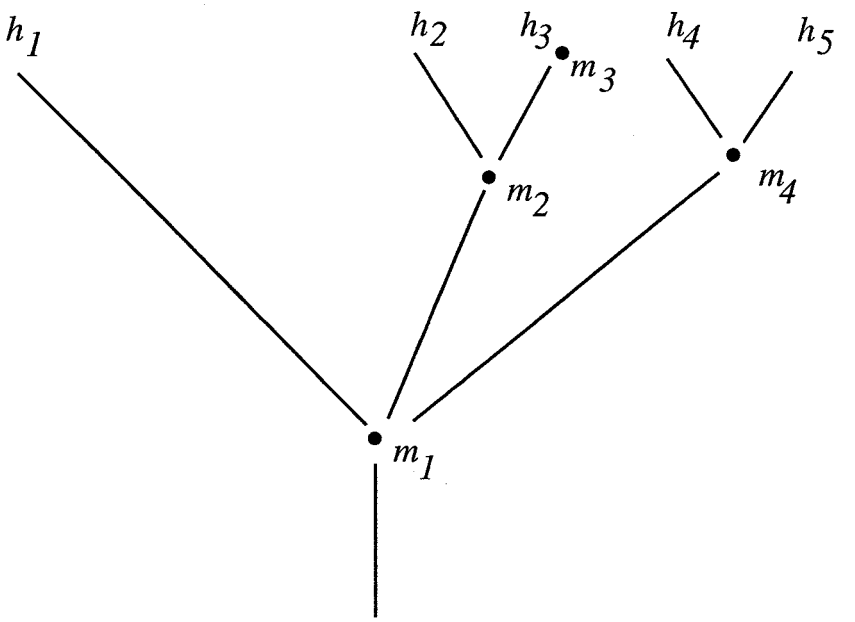


figure 6.1 logique temporelle indéterministe:
moments et histoires

Comme le note Horty, on peut trouver plus avantageux de prendre comme point de départ les histoires elles-mêmes structurées par leur type d'ordre : une relation d'ordre partiel strict[27]. Ayant défini la relation ≤, on peut définir ce type d'ordre < par la clause

$$m_1 < m_2 =_{\text{déf}} m_1 \leq m_2 \;\&\; m_1 \neq m_2$$

Pour les définitions qui suivent, nous suivons Horty et Belnap [1995][28]. Une *structure temporelle ramifiée* est un arbre $\mathcal{A}$ muni d'une relation d'ordre partiel strict $\langle \mathcal{A}, < \rangle$. Belnap désigne cet ordre temporel par l'expression « ordre causal strict » pour rappeler que les moments du temps ne sont pas des abstractions ou des représentations mais des événements concrets ; notons au passage que c'est le type d'ordre qui est causal et l'on ne doit pas interpréter

cette formule comme signifiant que les moments qui se suivent sont des causes et leurs successeurs des effets[29].

Dans la figure 6.1, la ligne verticale qui s'arrime au moment m_1 et qui constitue le tronc de l'arbre représente le passé. À partir du moment m_1 le futur se ramifie en diverses histoires qui forment les branches de l'arbre. Une histoire est un trajet complet qui part du pied de l'arbre et va jusqu'à un de ses sommets. Dans cette figure, cinq histoires différentes sont représentées (h_1,..., h_5). L'indéterminisme implique qu'un même moment peut être contenu dans un grand nombre d'histoires distinctes. L'ensemble des histoires qui passent par un moment m est noté $H_{(m)}$; la définition de cet ensemble est donnée par $H_{(m)} = \{h : m \in h\}$. Ainsi, dans la figure 6.1, $H(m_2) = \{h_2, h_3\}$.

On dit de deux histoires, par exemple h_2 et h_3, qu'elles sont indivises à un moment, par exemple au moment m_1 dans la figure, si, à ce moment, ces deux histoires sont indivises mais que ces histoires sont divisées à un moment ultérieur m_2. (Voir fig.6,1)

Un modèle pour la version de la logique du temps ramifié que nous venons d'esquisser est une structure $\mathcal{M} = < \mathcal{F}, v >$ où $\mathcal{F}$ est une structure temporelle et v est la fonction d'interprétation qui donne la valeur (dénotation) de chaque constante du langage propositionnel sous la forme d'un sous-ensemble de l'ensemble des paires m / h. Intuitivement, il s'agit pour chaque proposition de l'ensemble des points de l'index où cette proposition est vraie.

P est vraie au moment m de l'histoire h dans le modèle $\mathcal{M}$:

$\mathcal{M}, m / h \vDash P$ *ssi* $m / h \in v(P)$ où P est une formule atomique.

La conjonction :

$\mathcal{M}, m / h \models P \,\&\, Q$ *ssi* $M, m / h \models P$ et $M, m / h \models Q$

La négation :

$\mathcal{M}, m / h \models \neg P$ *ssi* $\mathcal{M}, m / h \not\models P$

On peut définir les modalités temporelles **P** et **F** sur la base de ces définitions.

Le passé :

$\mathcal{M}, m / h \models \mathbf{P}P$ *ssi* il y a $m' \in h$ tel que $m' < m$ et $\mathcal{M}, m' / h \models P$

Le futur :

$\mathcal{M}, m / h \models \mathbf{F}P$ *ssi* il y a $m' \in h$ tel que $m < m'$ et $\mathcal{M}, m' / h \models P$

La nécessité historique :

$\mathcal{M}, m / h \models \Box P$ *ssi* il y a $\mathcal{M}, m / h' \models P$ pour toute histoire $h' \in H_{(m)}$

Suivant l'usage, la possibilité historique sera exprimée par une règle d'abréviation $\Diamond P =_{\text{déf}} \neg \Box \neg P$. Il est utile de pouvoir exprimer dans le métalangage l'expression *la vérité de P est établie* à un moment dans un modèle si et seulement si $\mathcal{M}, m / h \models P$ pour chaque $h \in H_{(m)}$. Dans le même esprit, nous dirons que *la fausseté de P est établie* au moment m lorsque $\mathcal{M}, m / h \not\models P$ pour chaque $h \in H_{(m)}$.

Le concept de choix est défini comme une partition des histoires possibles. Plus précisément, on a une fonction de choix qui associe à chaque couple formé d'un agent α et d'un moment m une partition $Choix_{\alpha}^{m}$ des histoires $H_{(m)}$ qui passent par m. Les classes d'équivalence qui appartiennent à $Choix_{\alpha}^{m}$ représentent les choix possibles, dans le présent contexte, ce sont les actions qui sont des options pour α à m. Si K est une cellule de choix appartenant à $Choix_{\alpha}^{m}$, autrement dit, une des classes d'équivalence qui appartient à la partition, on dira que l'action K est une option pour α au

moment m. On dira que l'agent *accomplit* l'action K à m / h seulement si h appartient à K. On écrit $Choix^m_\alpha(h)$ (pour $h \in H_{(m)}$) pour décrire une action précise d'un agent α à m / h. Si deux moments m_1 et m_2 appartiennent à des histoires de la même partition $Choix^m_\alpha$, on dit que ces deux moments sont *équivalents relativement aux* $Choix^m_\alpha$.

La prochaine étape est de définir une structure d'interprétation possible pour la logique de l'action, une *structure stit* de la forme

$\langle Arbre, <, Agent, Choix, Instant \rangle$

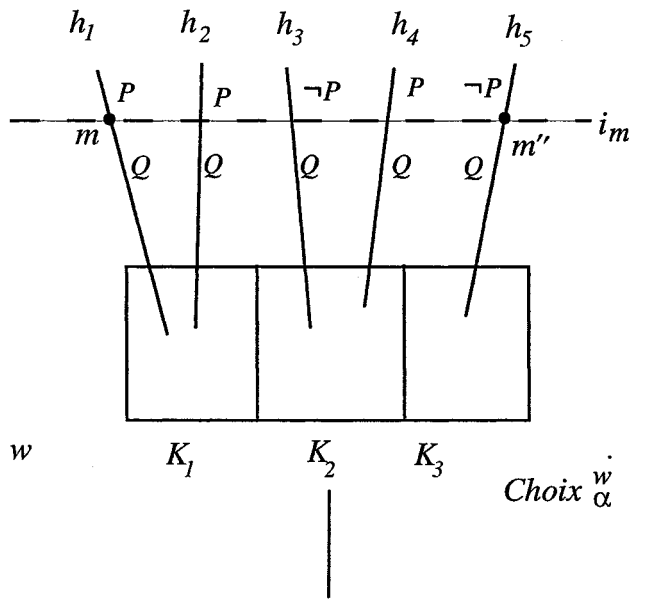


Figure 6.2 [α *astit*: P] est vrai à m/h_1

où *Arbre* et $<$ sont tels que définis antérieurement ; un modèle pour la logique de l'action *stit* est une structure $\mathcal{M} = < \mathcal{F}, \nu >$ où $\mathcal{F}$ est une structure stit et ν est une fonction d'interprétation qui donne la valeur (dénotation) de chaque

constante du langage propositionnel sous la forme d'un ensemble de paires *m / h.*

Comme nous l'avons suggéré au chapitre II, il est possible de représenter les résultats des actions d'un agent à un moment dans une histoire comme les actions *K* qui appartiennent à ce moment. Ces caractérisations sont minimalistes, mais elles rendent possible une explication plus précise de la notion d'action en ajoutant des clauses restrictives. Le passage du temps produit aussi assurément l'attrition des branches[30]. Il faut préciser davantage la contribution de l'agent. Considérons d'abord la forme logique de l'action comme accomplissement, l'opérateur *astit.* Nous pouvons formuler la règle pour évaluer une formule de la forme [α *astit* :*P*] à un point de référence *m / h* dans un modèle. Voici la définition de cette paraphrase *stit* pour l'expression α *fait en sorte que P* au sens de l'accomplissement.

α *astit* : *P*

$\mathcal{M}, m / h \models [\alpha\ astit : P]$ *ssi* il y a un moment $w < m$ tel que (1) pour tous les moments m' qui sont équivalents relativement au $Choix_{\alpha}^{w}$, nous avons $\mathcal{M}, m' / h' \models P$ pour tout h' tel que $h' \in H_{(m')}$; (2) Il y a un moment m'' $\in i_{(m)}$ tel que $w < m''$ et $\mathcal{M}, m'' / h'' \not\models P$ pour quelque $h'' \in H_{(m'')}$.

Dans la logique du *stit*, la clause (1) est appelée condition positive et elle indique qu'il est assuré maintenant — à *m* — que la proposition *P* sera vraie dans le futur en tant que résultat du choix de l'agent α au moment *w*, appelé le *témoin.* Le moment témoin est le moment du choix tandis que l'autre moment est celui de l'action, celui où *P* devient vrai. La clause (2) indique qu'il n'était pas fixé à *w* que *P* serait vraie maintenant — à $i_{(m)}$ — de telle sorte que l'action de l'agent y est pour quelque chose dans le fait que *P.*

Définition de la paraphrase *dstit* pour l'expression α *fait en sorte que P* au sens délibératif.

$\mathcal{M}, m / h \models [\alpha\ dstit : P]$ ssi (1) $\mathcal{M}, m / h' \models P$ pour chaque $h' \in Choix_{\alpha}^{m}(h)$, et (2) il y a au moins une histoire $h'' \in H_{(m)}$ où $\mathcal{M}, m / h'' \not\models P$.

Les conditions positives et négatives expriment des idées analogues à celles de la définition précédente. La condition positive est à l'effet que l'agent α doit agir au moment *m* de manière à assurer la vérité de *P*. La condition négative est à l'effet que la vérité de *P* ne soit pas déjà établie à tous les moments dans tous les futurs. Les différences entre les opérateurs *astit* et *dstit* sont les suivantes. Dans la définition de l'opérateur *astit*, on fait référence à deux moments du temps, le moment *m* de l'évaluation et le moment *w* qui sert de témoin. Le moment témoin est le moment du choix. Une seconde différence est que dans l'évaluation d'une formule comportant un *stit* d'accomplissement, *astit*, les histoires jouent un moindre rôle. Dans l'évaluation d'une formule comportant un opérateur *dstit*, les histoires jouent réellement un rôle. Alors que le *stit* d'accomplissement est entièrement tourné vers l'accomplissement de quelque chose par l'agent, le *stit* de délibération correspond à l'idée que quelque chose sera fait par suite d'un choix de l'agent.

Dans son article intitulé *Choice Trees*, McCall fut, à notre connaissance, le premier à tenter de situer la décision dans la représentation du temps ramifié sous forme d'arbre[31]. Comme il assigne au présent le caractère dynamique du flux temporel, il peut représenter la décision comme un pointeur qui indique un moment, exactement comme une flèche qui indique le « vous êtes ici » sur

une carte. Un tel pointeur indiquerait un nœud dans l'arbre représenté à la figure 6.1. Par contre, on ne voit pas ce que cette annotation ajouterait à la structure de l'arbre présenté à la figure 6.2.

Nous proposons la paraphrase suivante pour un concept d'action qui s'ajuste à cette notion de choix : une action de l'agent α à m émonde l'arbre des futurs possibles de façon à faire en sorte qu'une et une seule classe d'équivalence appartenant à la partition $Choix_{\alpha}^{m}$ soit réalisée. À ce niveau de généralité, nous pouvons feindre d'oublier l'aspect agentiel, à savoir que l'action de l'agent va faire en sorte que quelque chose se passe qui ne se serait pas passé autrement, exactement comme dans la situation décrite dans la figure 6.2. Notre paraphrase est équivalente à celle de Horty, qui considère qu'une action à un moment peut être vue simplement comme « le fait de contraindre le cours des événements futurs à se retrouver dans un sous-ensemble défini des histoires possibles qui sont disponibles à ce moment »[32].

Ni l'intentionnalité de l'action, ni la délibération, ni les probabilités ne sont intégrées à la théorie de base proposée par Belnap et Horty[33]. Cependant, il ne s'agit nullement d'un oubli ou d'un refus de reconnaître l'importance de l'intentionnalité de l'action. La position méthodologique de N. Belnap dans l'exposé de sa théorie des agents et des choix est énoncée clairement, en particulier pour ce qui est de l'importance de la délibération et de l'intentionnalité :

> Dans ce livre, nous n'offrons aucune idée qui pourrait aider à concevoir l'équipement mental ou quelqu'autre aspect de la « constitution interne réelle » des agents. La question est importante et elle est importante pour nous. C'est tout simplement que nos tentatives de progrès nous ont conduit dans une autre direction. Alan Ross Anderson avait l'habitude

> de dire que tout progrès en philosophie n'est possible qu'en supposant que certains problèmes ont été résolus alors qu'ils ne l'ont pas été et en poursuivant la recherche sous cette supposition.[34]

On peut s'arrêter pour considérer ce qui a été accompli. Nous disposons maintenant d'un langage qui tient compte de la dimension temporelle et qui permet d'exprimer l'idée d'acte et de résultat d'un acte d'une façon beaucoup plus détaillée et explicite que ce qu'on a pu rencontrer dans les diverses logiques de la décision, et ceci, de Ramsey à Joyce. Dans la logique temporelle que nous venons d'esquisser, nous disposons d'une locution beaucoup plus précise pour décrire le résultat d'une action ; ce n'est plus un monde, une proposition-monde, un état de chose, mais un ensemble d'histoires, un sous ensemble de l'ensemble des futurs possibles. Nous avons dit que les questions d'ontologie formelle — ce que Belnap appelle la métaphysique avec un petit « m » — sont complexes et qu'il est difficile de les aborder sans en fournir une formulation détaillée. Nous avons maintenant la base d'un langage précis pour le faire. Il est clair que les histoires sont des entités beaucoup plus finement discriminées, ce qui est en soi un avantage. On peut capitaliser sur cet avantage en cherchant à clarifier la notion de choix. Le problème classique de la théorie de l'action : « Qu'est-ce que l'agent fait au juste ? » a son équivalent dans la délibération : « Qu'est-ce que l'agent choisit au juste ? ». De plus, parce qu'une action est un engagement tourné vers le futur, cette approche de l'action correspond mieux à nos intuitions à propos de la nature de la délibération et de l'action. Dans la prochaine section, nous allons considérer la délibération et la décision du point de vue des traits

fondamentaux de la logique de l'action et des agents situés dans un monde indéterministe.

6.3. Infrastructure logique, délibération et décision

Dans cette section, nous allons passer en revue quelques composantes de la délibération que nous nous sommes engagés à décrire dans les chapitres précédents, incluant ceux pour lesquels la logique de l'action et du temps ramifié offre un langage approprié. Ce faisant, nous ne prétendons pas avoir trouvé un moyen de donner une représentation abstraite perspicace de la délibération mais nous cherchons simplement à mettre en lumière ce qu'une théorie de la décision devrait pouvoir développer. C'est pourquoi la section est ordonnée comme une énumération d'éléments accompagnés de commentaires. Le temps disponible ne nous a pas permis d'exposer nos idées de façon plus détaillée.

Arbre de décision et arbre du temps ramifié

Il est remarquable que la logique de l'action dans la structure du temps ramifié reprenne une représentation de l'action et du temps qui fut proposée par von Neumann et Morgenstern pour représenter le développement d'un jeu en forme stratégique[35]. L'analyse des problèmes de décision représente souvent ces problèmes comme des arbres où les noeuds qui correspondent à des choix (l'agent décide) sont distingués des nœuds qui correspondent à la chance (la nature décide). Les arbres de la logique temporelle et les arbres de décision entretiennent des rapports réels, mais peu décrits avec les arbres de

choix[36]. Dans ce qui suit, nous supposerons qu'un arbre de décision est une représentation comparable à une petite carte géographique ou au plan d'une ville représentant une portion limitée d'un arbre au sens de la logique du temps ramifié[37]. Les éléments représentés sur une carte ont une structure discrète, pas nécessairement l'arbre du temps ramifié[38]. Si on représente le choix dans un arbre de décision, alors à chaque nœud, un seul agent agit. Ceci n'est pas présupposé dans l'arbre du temps ramifié. Sans prétendre avoir épuisé le sujet, nous concluons ce commentaire en soulignant que ceux qui s'intéressent à l'infrastructure logique du choix en théorie de la décision devraient s'inquiéter davantage de cette question[39].

L'arbre en tant que partition

Une propriété intéressante des arbres du point de vue de la décision est que la ramification à partir du moment présent en une série d'histoires correspond très exactement à l'idée de partition. Comme le disait von Neumann, une partition est un système de constructions qui couvrent exhaustivement l'ensemble des possibilités et qui n'est pas absurde[40]. Autrement dit une collection d'histoires disjointes deux à deux qui épuise les possibles (ici, les futurs possibles) à partir de l'instant du choix. C'est sur un tel ensemble que doivent se répartir les degrés de croyance de l'agent pour être représentables par une fonction de probabilité[41]. L'arbre nous donne l'occasion de retrouver ce qui est fondamental dans l'idée de partition. Au chapitre V, nous avons vu qu'il y bien des débats d'interprétation en théorie causale de la décision sur la nature exacte des partitions.

On veut représenter la totalité de ce qui importe pour l'agent confronté à un choix dans l'incertitude (l'arbre de décision) parce qu'il habite notre monde qui est indéterministe (l'arbre du temps ramifié). Dans un arbre de décision, une partition est une déclaration préliminaire qui indique les différentes directions dans lesquelles seront précisées des informations ultérieurement. Dans l'arbre du temps ramifié c'est le passage du temps qui va garantir l'émondage des branches de l'arbre et le rétrécissement de la classe des futurs. Tout ceci correspond d'assez près à la nature du choix tel que décrit par la théorie de la décision et par la construction $Choix_{\alpha}^{m}$. Il s'agit d'une combinaison comportant une part de détermination (je choisis X et non pas Y) qui contient aussi une part d'indétermination (je ne sais pas encore tout ce que contient X). Cette façon de formuler les choses donne autant de précision que possible et cette mixture de détermination et d'indétermination est liée à l'essence même de la notion de choix.

Le choix et son résultat

Nous avons dit que les propositions atomiques devraient trouver leur place dans des constructions logiques plus complexes représentant des faits momentanés actuels qui se produisent par suite d'actions des agents. Nous pouvons maintenant en dire davantage sur la nature de ces constructions. Nous affirmons que les constructions logiques qui représentent des familles de futurs possibles sont de meilleures représentations des objets de choix que tous les objets utilisés pour jouer ce rôle dans les théories de la décision proposées à ce jour. L'expression qui est proposée par Horty et Belnap

« contraindre le cours des événements futurs à se retrouver dans un sous-ensemble défini des histoires possibles » traduit mieux nos intuitions quant à la nature du choix. L'importance de l'idée de tentative dans l'analyse de la délibération, importance que nous avons eu l'occasion de souligner en discutant la théorie de Jeffrey, peut être clarifiée en logique de l'action. Daniel Vanderveken a montré la direction à suivre en définissant des opérateurs modaux pour représenter les tentatives dans une logique de l'action comparable à celles que nous avons envisagées. Ultérieurement, il nous appartiendrait de montrer qu'on peut vraiment tirer profit de cette formulation pour exprimer le modèle de la délibération dans le cadre de la formulation d'une théorie causale de la décision[42].

La décision

Le contexte de la logique de l'action et du temps ramifié que nous avons décrit plus haut rend possible une discussion de l'idée même de décision. Lorsqu'on tente de définir un concept comme celui de décision, il est utile de comprendre sa grammaire logique, c'est-à-dire principalement la façon dont il s'arrime à d'autres termes de base utilisés pour décrire la délibération, comme les opérateurs *stit* par exemple. Avant d'examiner la syntaxe de cet opérateur, nous devons d'abord clarifier nos idées sur sa sémantique. Quelles pourraient être les conditions de vérité d'une décision de faire *P* ? Nous cherchons un bon candidat pour compléter la forme « Jones a décidé de beurrer la tartine si et seulement si.... ». Certains ont été tenté d'utiliser une croyance de l'agent pour compléter la formule[43]. Si Jones a décidé de faire *P*, Jones *sait* qu'il a décidé de faire *P* et réciproquement, lorsqu'un agent *sait* qu'il a décidé de

faire *P*, c'est qu'il a décidé de le faire. L'équivalence stricte des deux conditions ne s'expose à aucune critique que nous puissions imaginer. Cette clause pourrait donc convenir. La croyance qui découle de la condition d'introspection remplit toutes les conditions requises. Si Jones a décidé de faire *P*, il le sait et s'il le sait, c'est qu'il a décidé. Cette sémantique de l'opération de décision serait tout à fait acceptable, mais selon nous, elle a une conséquence indésirable. Elle explique la décision comme une condition épistémique de l'agent plutôt que comme une véritable action. Il y a là une question importante qui relève de la philosophie de l'esprit. Les actions ont des conditions de succès — parfois des conditions de satisfaction — qui leurs sont propres et c'est une tâche pour l'analyse logique de la délibération et de la décision de les exposer explicitement.

Selon nous, il serait préférable de considérer la décision et l'acte de choisir comme des actes mentaux[44] de façon à marquer leur continuité avec les autres actes, et en particulier avec la formation d'une intention[45]. Pour nous, les décisions sont de véritables actions et non pas simplement des états doxastiques de l'agent. À l'appui de notre position, nous soumettons qu'une décision a toutes les propriétés d'une action ordinaire dans l'environnement de l'opérateur *dstit* et sa logique propre. Elle vérifie tous les slogans qui caractérisent cette conception de l'action. En effet, pour que la décision soit prise, il faut (1) que quelque chose se passe. De plus, (2) une décision n'est jamais prise avant son temps. Ce qui n'empêche pas que l'on puisse décider de reporter une décision, exactement comme on le fait régulièrement pour d'autres actions. (3) Nous ne pouvons pas prendre la décision de faire quelque chose qui se produirait nécessairement sans que nous le décidions. Enfin, (4) quand une décision est prise, il est vrai qu'elle a été prise, et ce, bien qu'elle

soit révisable. Nous croyons aussi qu'il est possible de prédire que nous allons prendre une décision, de même que nous pouvons prédire nos propres actions[46]. Ces observations ne nous donnent pas une conception complète des actes de délibération et de décision mais ces éléments sont suffisants pour constituer les bases d'une véritable analyse logique de la délibération et de la décision. Le terme « décider » a un sens principal que nous tentons de cerner et plusieurs sens dérivés. Ainsi nous proposons de remplacer l'interprétation « décider, c'est croire que l'on a décidé » par la paraphrase « décider, c'est se résoudre à tenter ». Bien entendu, nous ne soutenons pas que la décision soit réductible à un simple acte quelconque, un acte « tout court ». Notre suggestion est à l'effet que tout langage qui possède un opérateur pour les verbes intentionnels « vouloir » et « tenter » pourrait inclure un opérateur pour le verbe « décider » dont la sémantique peut être précisée dans l'esprit des observations qui précèdent.

Considérons maintenant les questions de grammaire. L'idée d'inclure un opérateur de décision dans le langage-objet nous est venue d'un passage dans Joyce[1999] où l'auteur utilise un tel opérateur comme un élément du métalangage. Il utilise cet opérateur pour formuler la définition d'une condition qui fait écran à une autre, condition que nous avons expliquée à la section 6.1.[47] S'il apparaissait utile d'introduire un opérateur de décision, nous croyons qu'un opérateur tel opérateur de décision pourrait s'inscrire dans la forme $\alpha\mathbf{d}[\alpha\ dstit : P]$. Cette forme est l'abréviation de l'expression « l'agent décide qu'il fera en sorte que P ». À notre connaissance, les possibilités offertes par cette analyse n'ont pas encore été explorées. Cependant, il est possible que la logique de l'opérateur **d** ait déjà fait l'objet de recherches logiques dans un autre contexte formel que ceux avec lesquels

nous sommes familiers. En conformité avec l'esprit exploratoire annoncé pour cette section, nous terminons en proposant quelques évidences à propos de l'opérateur **d** qui doivent être comparées à des formules analogues de la logique des opérateurs de l'action :

$\alpha\mathbf{d}\,[\alpha\ dstit : P] \supset \neg\,\alpha\mathbf{d}\,[\alpha\ dstit : \neg P\,]$

Un agent a décidé de faire en sorte que *P* seulement si cet agent ne prend pas la décision de faire en sorte que ¬*P*. Cette formule sera valide dans les modèles où chaque décision a son propre moment. En indexant cette formule relativement à des intervalles de temps, on pourrait formuler les principes élémentaires de la cohérence dynamique dans la délibération. On note en passant que la réciproque de cette formule n'est pas valide car le conséquent négatif $\neg\,\alpha\mathbf{d}\,[\alpha dstit : \neg P\,]$ n'implique aucune décision de la part de l'agent. La formule suivante encode l'observation que le contenu d'une décision ne se réalise pas en conséquence de la décision est prise :

$\neg\,[\alpha\,\mathbf{d}\,[\alpha\ dstit : P\ ] \supset P]$

En d'autres termes, certaines décisions d'agir ne sont pas suivies par un moment où *P* est établie vraie.

$\alpha\mathbf{d}\,[\alpha dstit : P]\ \&\ \alpha\mathbf{d}\,[\alpha dstit : Q] \supset\ \alpha\mathbf{d}\,[\alpha dstit : P\ \&\ Q]$

Pour être valide, cette formule doit satisfaire la condition que *P et* Q sont compatibles.

Enfin, la formule suivante n'est pas valide dans les modèles où chaque choix a son propre moment, sous peine de régression à l'infini :

$\alpha\mathbf{d}\,[\alpha dstit : P] \supset \alpha\mathbf{d}\ [\alpha\mathbf{d}\,[\alpha dstit : P]]$

BIBLIOGRAPHIE

La méthode utilisée dans le corps de la thèse pour référer aux entrées de cette bibliographie dérive de celle proposée et expliquée à l'origine par S. C. Kleene dans *Introduction to metamathematics,* Wolters-Noordhoff et North-Holland, 1re édition, 1952. Elle est couramment utilisée en mathématiques et en sciences humaines. L'expression formée d'un nom propre immédiatement suivi d'une année apparaissant entre crochets constitue une référence à cette bibliographie. Dans le corps du texte, une telle expression est utilisée comme le nom d'un article ou d'un ouvrage. L'année de la première publication est prise comme point de référence et, en général, c'est elle qui est utilisée. Lorsqu'il existe des éditions successives comportant des révisions, l'année de l'édition qui a été consultée, si elle est ultérieure, est mentionnée dans l'entrée bibliographique de façon à valider les références aux numéros des pages citées. Dans le cas de publications posthumes, il a semblé légitime d'utiliser l'année de rédaction du texte lorsqu'elle était connue. Ceci permet d'aider le lecteur à mieux situer la chronologie des publications citées. Toujours pour valider les références aux numéros de page, lorsqu'un article a été reproduit dans une autre publication, par exemple un recueil, l'édition utilisée est mentionnée à son entrée dans cette bibliographie. Enfin, conformément à l'usage prépondérant, lorsque plusieurs écrits d'un même auteur sont associés à une même année, les lettres « a », « b », « c », sont utilisées pour les distinguer.

En général, les ouvrages cités dans cette bibliographie sont ceux qui sont cités dans le corps du texte, dans les notes ou encore à d'autres entrées de cette bibliographie. Nous avons cherché à restreindre le nombre de titres correspondant à des ouvrages consultés dans la préparation de la thèse mais qui ne sont pas cités dans le corps de la thèse ou dans les notes.

ABDELLAOUI, M. et WAKKER, P. P.

[2004] « A Uncertainty-Oriented approach to Subjective Expected Utility and its Extensions », 11th Conference on Foundations of Utility and Risk Theory, (FUR'04) GRID, Maison de la Recherche de ESTP,

Cachan, France. Texte disponible (09/04) à l'adresse suivante : http://www1.fee.uva.nl/creed/wakker/.

ALLAIS, M.

[1953] « Le comportement de l'homme rationnel devant le risque : Critiques des postulats et axiomes de l'école américaine », *Econometrica*, 21, p. 503-546.

[1979] « Criticism of the Postulates and Axioms of the American School », dans Allais, M. et Hagen O. dirs. [1979] ; traduction et version revue de Allais [1953], reproduite dans Moser [1990], p. 113-140.

ALLAIS, M. ET HAGEN O, (dirs.)

[1979] *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, D. Reidel, 1979, p. 67-95.

ANSCOMBE, G. E. M.

[1957] *Intention*, Blackwell, Oxford ; réimpr. Cornell University Press, NY, 1966.

ARISTOTE

[-/1990] *Éthique à Nicomaque* ; traduit et édité par J. Tricot, Librairie Philosophique J.Vrin.

ARMENDT, B.

[1986] « A Foundation for Causal Decision Theory », *Topoi*, 5, p. 3-19.

ARNAULD & NICOLE

[1680] *La logique ou l'art de penser* ; édité avec une introduction de Louis Marin, coll. Champs, Flammarion, Paris (1970).

ARROW, K.J., COLOMBATTO, E., PERLMAN, M., SCHMIDT, C. (éds.)

[1996] *The Rational Foundations of Economic Behavior*, Actes de la conférence de IEA tenue à Turin, Italie, New York, St-Martin's.

AUDI, R. (dir.)

[1999], *The Cambridge Dictionary of Philosophy*, 2e édition, Cambridge, Cambridge University Press.

AUER, L. VON

[1999] « Dynamic choice mechanisms », *Theory and Decision*, 46, p. 291-308.

AUMANN, R. J. et ANSCOMBE, F.J.

[1963] « A Definition of Subjective Probability », *Annals of Mathematical Statistics*, vol. 34, p. 199-205.

BACHARACH, M. O. L., GERARD-VARET, L. A., MONGIN, P., et SHIN, H.S.

[1997] *Epistemic Logic and the Theory of Games and Decisions*, Theory and decision library, Series C, Volume 20, Dordrecht, Kluwer Academic Publishers.

BACHARACH, M., HURLEY, S.

[1991] *Foundations of Decision Theory*, Oxford, Basil Blackwell.

BARBERA, S., HAMMOND, P. et SEIDL, C. (dirs.)

[2004] *Handbook of Utility Theory*, 2 tomes, Kluwer Academic Press.

BAYES, T.

[1763] « An Essay toward solving a problem in the doctrine of chances », *Philosophical Transactions of the Royal Society*, 53, p. 370-418 ; nouvelle édition publiée dans *Biometrika*, 45, (1958), p. 296-315.

BELL, D. E., RAIFFA, H. et TVERSKY, A, (dirs.),

[1988] *Decision Making: Descriptive, Normative and Prescriptive interactions,* Cambridge, Cambridge University Press.

BELL, J. L.

[1998] *A Primer of Infinitesimal Analysis*, Cambridge University Press.

[2000] « Continuity and the Logic of Perception », *Transcendent Philosophy*, vol.1, no. 2.

[2005] ? « An invitation to smooth infinitesimal analysis » ; à paraître, dans Harper, W. et Myrvold, W., dirs., *Infinitesimals : Concepts and Applications*, Western Ontario Series in the Philosophy of Science, Kluwer.

BELNAP, N.

[1966] « Comment on H. Simon's paper, « The Logic of Heuristic Decision Making », p. 27-31, dans Rescher [1966] *The logic of decision and action.* Pittsburgh, University of Pittsburgh Press.

[1994] *The Stit Papers*, Pittsburgh, department of philosophy ; recueil d'articles polycopiés dont certains sont rédigés en collaboration ; il vaut mieux consulter Belnap, N. Perloff, M. et Xu, M. [2001] qui a un contenu comparable aux articles de ce recueil.

BELNAP, N. et PERLOFF, M.

[1990] « Seing to it that : a canonical form for agentives » dans H. E. Kyburg, jr., R.P. Loui et G.N. Carlson (dirs.) *Knowledge, representation and defeasible reasoning*, Kluwer Academic, p. 175-199 ; document 1 de Belnap [1994].

BELNAP, N. PERLOFF, M. ET XU, M.

[2001] *Facing the future : Agents and Choices in Our Indeterminist World,* New York, Oxford University Press.

BERNOULLI, D.

[1738] « Specimen Theoriae Novae de Mensura Sortis », *Commentarii academiae scientiarum imperialis Petropolitanae*, 5, p. 175-192 ; trad. angl. L. Sommer, « Exposition of a new theory of the measurement of risk », *Econometrica*, 22, 23-36, 1954.

BICCHIERI, C.

[1993] *Rationality and Coordination,* Cambridge, Cambridge Studies in Probablility, Induction, and Decision Theory, Cambridge University Press.

BICCHIERI, C., JEFFREY, R., SKYRMS, B., (dirs.)

[1999] *The Logic of Strategy,* Cambridge, Oxford University Press.

BLUME, L. BRANDENBURGER, A., et DEKEL, E.

[1991] « Lexicographic Probabilities and Choice Under Uncertainty », *Econometrica*, Vol. 59, no. 1, p. 61-79.

BODEN, M. A.

[1990] *The Philosophy of Artificial Intelligence,* New York, Oxford University Press.

BOLKER, E.

[1966] « Functions resembling quotients of measures » *Transactions of the American Mathematical Society*, vol. 124, p. 292-312.

[1967] « A Simultaneous Axiomatization of Utility and Subjective Probability », *Philosophy of science*, 34, p. 333-340.

BOREL, É.

[1924] « À propos d'un traité des probabilités », *Revue philosophique*, 98, p.321-336, trad. angl. dans Kyburg et Smokler [1964] dirs., p 47-60.

BOUVIER, A. et GEORGE, M.

[1979] *Dictionnaire des Mathématiques*, Paris, Presses Universitaires de France ; sous la direction de F. Le Lionnais, 2e édition revue et mise à jour.

BRADLEY, R.

[1999] « Conditional Desirability », *Theory and Decision*, 47, p. 23-55.

[2001] « Ramsey and the Measurement of Belief », dans Corfield, D. et Williamson, J. [2001] p. 273-299.

[2001]b « Review of *The foundations of causal decision theory* by James Joyce », voir Joyce [1999], *Economics and Philosophy*, vol. 17, no. 2, 281-288.

[2003] « Axiomatic Bayesian Utilitarianism », Cahier #2003-09, Laboratoire d'économétrie, École Polytechnique, CNRS, 29 p.

[2003b] « Probabilism and Mental Kinematics » Draft paper, Bayesian Epistemology Workshop at the 26th Wiitgenstein Colloquium, Kirchberg, Austria.

BRATMAN, M. E.

[1987] *Intention, Plans and Practical Reason*, Cambridge, Harvard University Press.

[1999] *Faces of Intention : Selected Essays on Intention and Agency*, Cambridge, Cambridge University Press.

BROOME, J.

[1990] « Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism », *Review of Economic Studies*, 57, p. 477-502.

BYRON, M., (dir.)

[2004] *Satisficing and Maximizing : Moral Theorists on Practical Reason*, Cambridge, Cambridge University Press.

CAMPBELL, R. et SOWDEN L., (dirs.)

[1985] *Paradoxes of Rationality and Cooperation*, Vancouver, UBC Press.

CARNAP, R.

[1937] *The Logical Syntax of Language*, International Library of Psychology, Philosophy and Scientific Method, London, Routledge and Kegan Paul.

[1945] « On Inductive Logic », *Philosophy of science*, 12, p. 72-97.

[1947] « On the Application of Inductive Logic », *Philosophy and Phenomenological Research* , 8, p. 133-147.

[1950-4] *Logical Foundations of Probability,* Chicago, University of Chicago Press.

[1952] *The Continuum of Inductive Methods,* Chicago, University of Chicago Press.

[1955] « Statistical and Inductive Probability », fascicule publié par *The Galois Institute of Mathematics and Art*, Brooklyn, N.Y. reproduit dans Madden, Edward H. *The Structure of Scientific Thought: An Introduction to the Philosophy of Science,* Boston, H. Miffin, 1960, p. 269-279.

[1962a] « The Aim of Inductive Logic », dans Ernest Nagel, Patrick Suppes, and Alfred Tarski, *Logic, Methodology and Philosophy of Science*. Stanford, Stanford University Press. (1962)

[1962b] *Logical Foundations of Probability, preface to the second edition of* Carnap, R. [1950-4].

[1963a] « Intellectual Autobiography », dans Schilpp P. A. [1963].

[1963b] « My Basic Conceptions of Probability » section V de « Replies and Systematic expositions », dans Schilpp, P.A. [1963], p. 966 - 973.

[1980] « A Basic System of Inductive Logic, Part 2 », dans Carnap et Jeffrey [1980].

CARNAP, R. et JEFFREY, R., (dirs.)

[1971] vol. 1 et [1980], vol. 2, *Studies in Inductive Logic and Probability*, Berkeley, University of California Press.

CHANG, C. C. et KEISLER H. J.

[1973] *Model Theory*, Studies in Logic and The Foundations of Mathematics, Vol. 73, North-Holland ; 3e édition (1990).

CHAPUIS, A.

[2000] « Rationality and Circularity », dans Chapuis et Gupta [2000], p. 49-78.

CHAPUIS, A. et GUPTA, A. (dirs.)

[2000] *Circularity, Definition and Truth*, Indian Council of Philosophical Research, Munshiram Manoharlal Publishers, New Delhi.

CHELLAS, B. F.

[1980] *Modal Logic : An Introduction*, Cambridge, Cambridge University Press.

CHERNIAK, C.

[1986] *Minimal Rationality*, Cambridge, MIT Press, Bradford books.

CHERNOFF, H. et MOSES, L. E.

[1959] *Elementary Decision Theory,* New York, Wiley, et Toronto, Dover (1986).

CHINSTEIN, P. A. et HANNAWAY, O., (dirs.)

[1985] *Observation, Experiment, and Hypothesis in modern Physical Science,* Cambridge, MIT Press.

CHISHOLM, R.

[1979] « On the Logic of Purpose », dans French, P. A., Uehling, T. E. et Wettstein, H. K., dirs., Midwest studies in philosophy, Vol IV, Minneapolis, Univ. of Minnesota Press, p. 223-237.

CONNOLY, T., ARKES, H. R., HAMMOND, K. R., (dirs.)

[2000] *Judgment and Decision Making : An interdisciplinary Reader*, 2e édition, Cambridge Series on Judgment and Decision Making, Cambridge, Cambridge University Press.

CORFIELD, D. et WILLIAMSON, J (dirs.)

[2001] *Foudations of Bayesianism*, Dordrecht, Kluwer Academic Press

CUMMINS, R. et POLLOCK, J.

[1991] *Philosophy and AI : Essays at the Interface*, Cambridge, MIT Press, Bradford Books.

DASTANI, M., DIGNUM, F., MEYER, J.-J.

[2003] « Autonomy and Agent Deliberation », Proceedings of The First International Workshop on Computatinal Autonomy - Potential, Risks, Solutions (Autonomous 2003), Michael Rovatsos and Matthias Nickles (éds.), 23-35, Melbourne, juillet 2003.

DAVIDSON, D.

[1963] « Actions, Reasons, and Causes », Essay 1 de Davidson [1980].

[1974] « Psychology as Philosophy », Chapitre 12 de Davidson [1980] ; ce texte a d'abord été publié dans un ouvrage collectif sous la direction de S. C. Brown intitulé *Philosophy of Psychology*, Macmillan Press and Barnes, Noble, Inc., 1974.

[1976] « Hempel on Explaining Action », Essay 14 de Davidson [1980].

[1980] *Essays on Actions and Events*, Oxford, Oxford University Press ; nous citons l'édition revue et corrigée qui date de 1982.

DAVIDSON, D. et SUPPES, P.

[1956] « A Finististic Axiomatization of Subjective Probability and Utility », *Econometrica*, 24, p. 264-275.

DAVIDSON, D., SUPPES, P. et SIEGEL, S. (dirs.)

[1957] *Decision making*, Stanford, Stanford University Press.

DEBREU, G.

[1959] « Cardinal Utility for Even-Chance Mixtures of Pairs of Sure Prospects », *Review of Economic Studies*, 26, p. 174-177.

DE FINETTI, B.

[1936] « La Logique de la Probabilité », Actualités scientifiques et industrielles, 391, *Actes du Congrès International de Philosophie Scientifique,* Sorbonne, Paris, 1935, IV, Induction et Probabilité, Herman et Cie, Éditeurs, Paris, 1936, 31-39. Cité selon la traduction anglaise de R. B. Angell « The Logic of Probability », *Philosophical studies*, 77, p. 181-190, 1995.

[1937] « La Prévision, ses Lois Logiques, ses Sources Subjectives », *Annales de l'Institut Henri Poincaré*, 7, p. 1-68 ; trad. anglaise sous le titre « Foresight : Its Logical Laws, Its Subjective Sources », dans Kyburg, H., et Smokler, H. [1964], p. 93-158.

DOMOTOR, Z.

[1978] « Axiomatization of Jeffrey's Utilities », Synthese, 39, p. 165-210.

DREZE, J. et RUSTICHINI, A.

[2004] « State-dependent Utility Theory », chap. 16 de Barbera, Hammond et Seidl [2004]. (ce texte était disponible sur le site web de Jacques Drèze)

DUBUCS, J.-P., (dir.),

[1993] *Philosophy of Probability*, Dordrecht, Philosophical Studies Series, Vol. 56, Kluwer Academic Publishers.

DUNN, J. M. and GUPTA, A. (dirs.),

[1990] *Truth or Consequences : Essays in honor of Nuel Belnap*, Kluwer Academic Publishers,

DUPUY, J.-P. et LIVET, P. (dirs.)

[1997] *Les limites de la rationalité,* Tome 1: *Rationalité, Éthique et Cognition,* Paris, La Découverte.

EARMAN, J.

[1992] *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory,* Cambridge, MIT Press.

EELLS, E.

[1982] *Rational Decisions and Causality,* Cambridge, Cambridge University Press.

[1984] « Metatickles and the dynamics of deliberation », *Theory and decision*, 17, p. 71-95.

[1985] « Causality, Decision and Newcomb's paradox » dans Campbell et Sowden [1985], p. 183-213.

[1999] « Prediction, Probability, and Pragmatics », dans *Canadian Journal of Philosophy,* Vol. 30, (2000), No. 2, p. 183-206.

[2000] « Review of *The foundations of causal decision theory* by James Joyce », *British Journal for the Philosophy of Science*, 51, p. 893-900.

EELLS, E. et SKYRMS, B.

[1994] *Rational Decisions and Causality,* Cambridge Studies in Probability, Induction and Decision Theory, New York et Cambridge : Cambridge University Press.

[1996] *Probability and Conditionals*, New York et Cambridge : Cambridge University Press.

ELLSBERG, D.

[1961] « Risk, ambiguity and the Savage axioms », *Quaterly Journal of Economics*, 75 (1961), p. 643-669 ; reproduit dans Gardenfors et Sahlin [1988], p. 245-269.

ELSTER, J.

[1979] *Ulysses and the Sirens, Studies in Rationality and Irrationality,* Éditions de la maison des sciences de l'homme et Cambridge, Cambridge University Press.

[1983] *Sour Grapes: Studies in the Subversions of Rationality,* Éditions de la maison des sciences de l'homme et Cambridge, Cambridge University Press.

[1989] *Solomonic Judgements: Studies in the Limitations of Rationality,* Éditions de la maison des sciences de l'homme et Cambridge, Cambridge University Press.

FAGIN, R. et HALPERN, J. Y.

[1994]« Reasoning about Knowledge and Probability », *Journal of the ACM*, Vol. 41, No. 2, p. 340-367.

FAGIN, R., HALPERN, J.Y., MOSES, Y. et VARDI, M.Y.

[1995] *Reasoning about Knowledge*, Cambridge, Mass, MIT Press.

FETZER, J. H.

[1988] *Probability and Causality,* Dordrecht, D. Reidel.

FISHBURN, P. C.

[1964] *Decision and Value Theory*, New York : Wiley.

[1970] *Les Mathématiques de la Décision*, traduit de l'anglais par Elliot Cohen, collection Mathématiques et Sciences de l'Homme, Vol. XVII, Mouton, Gauthier-Villars.

[1981] « Subjective Expected Utility : A Review of Normative Theories », *Theory and Decision*, 13, p. 139-199.

[1988] « Normative Theories of Decision Making Under Risk and Uncertainty », chapitre 4 de Bell, D. E. , Raiffa, H. and Tversky, A. [1988], p. 78-98.

[1994] « Tales of a Radical Bayesian », compte rendu de Jeffrey [1992], *Journal of Mathematical Psychology*, 38, p. 135-146.

FITELSON, B.

[2003] « Review of *The Foundations of Causal Decision Theory* by James Joyce », *Mind*, 112, no. 447, p. 545–551.

FRENCH, P. A., UEHLING, T. E., et WETTSTEIN, H. K., (dirs.)

[1990] *The Philosophy of the Social Sciences,* Midwest Studies in Philosophy, Volume XV, Notre Dame, University of Notre Dame Press.

FREUND, J. E.

[1973] *Introduction to Probability,* New York, Dover Publications.

GALAVOTTI, M. C.

[1991] « The notion of subjective probability in the work of Ramsey and de Finetti », *Theoria* , Vol. 57, p. 239-259.

[2005] *Philosophical Introduction to Probability*, CSLI lecture notes no 167, Stanford.

GARDENFORS, P.

[1984] « The Dynamics of Belief as a Basis for Logic », *British Journal for the Philosophy of Science*, Vol. 35, #1.

[1988] *Knowledge in flux: Modeling the Dynamics of Epistemic States*, Cambridge, MIT Press, Bradford Books.

GARDENFORS, P. et SAHLIN, N.-E. (dirs.)

[1988] *Decision, Probability, and Utility*, Cambridge, Cambridge University Press.

GARDNER, M.

[1973] « Free will revisited, with a mind-bending prediction paradox by William Newcomb », *Scientific American*, July1973, p. 104-109.

GAREY, M.

[1979] *Computers and Intractability : A Guide to the Theory of NP-Completeness*, Bell Telephone Laboratories, W. H. Freeman.

GAUTHIER, D.

[1985] « The Unity of Reason : A subversive Reinterpretation of Kant », chap. 5 de Gauthier [1990a], p. 110-126.

[1990]« Reason and Maximization », chap. 10 de Gauthier [1990a]; Ce texte est d'abord paru dans le *Canadian Journal of Philosophy*, 4, 1975, p. 463-471.

[1990]a *Moral Dealing : Contract, Ethics and Reason*, Cornell University Press, New York.

[1996] « Resolute Choice and Rational Deliberation : A Critique and Defence ». Cahiers d'Épistémologie, Université du Québec à Montréal, no 9611.

GIBBARD, A. et HARPER, W.

[1978] « Counterfactuals and Two Kinds of Expected Utility », dans Hooker, C. A., Leach, J. J. et Mc Lennen E. F. [1978], (dirs.) ; reproduit dans Gardenfors et Sahlin [1988], dir., chap. 17, p. 341-376.

GIGERENZER, G. et SELTEN, R.

[2001] *Bounded Rationality : The Adaptive Toolbox*, Dahlem Workshop Reports, Cambridge et Londres, The MIT Press.

GUPTA, A.

[2000] « On Circular concepts » dans André Chapuis and Anil Gupta, *Circularity, Definition and Truth*, Indian Council of Philosophical Research, p. 23-153.

HACKING, I.

[1975] *The Emergence of Probability*, Cambridge, Cambridge University Press.

HÁJEK, A.

[2001] « The Reference Class Problem is Your Problem Too », ; version xerox de la conférence SEP 2001 à Montreal, à paraître dans *Synthese* en 2005.

[2003] « What Conditional Probability Could Not Be », *Synthese*, Vol. 137, No. 3, December 2003, 273-323.

HALMOS, P.

[1974] *Lectures on Boolean Algebras*, Springer-Verlag ; première édition Van Nostrand (1963).

HALPERN, J. Y. et CHU, F. C.

[2003] « Great expectations. Part I: On the customizability of generalized expected utility », *Proceedings of the 18th International Joint Conference on Artificial Intelligence* (IJCAI 2003), p. 291-296.

[2003]b « Great expectations. Part II: Generalized expected utility as a universal decision rule » *Proceedings of the 18th International Joint Conference on Artificial Intelligence* (IJCAI 2003), p. 297-302.

HALPERN, J. Y. et PUCELLA, R.

[2002] « Reasoning about expectation », *Proceedings of the Eighteenth Conference on Uncertainty in AI*, p. 207-215.

HARPER, W. et HOOKER, C. (éds.)

[1976] *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Vol. 3., Dordrecht, Reidel.

HARGREAVES HEAP, M., HOLLIS, M., LYONS, B. SUGDEN, R. et WEALE, A.

[1992] *The theory of choice : A critical Guide*, Blackwell, Oxford.

HARSANYI, J. C.

[1975] « Nonlinear Social Welfare Functions », *Theory and Decision*, 7, p. 61-82.

HAMMOND, P. J.

[1999] « Consequentialism, Non-Archimedean Probabilities, and Lexicographic Expected Utility », chap. 2 de Bicchieri, C., Jeffrey, R., Skyrms, B. [1999], p. 39-66.

HARMAN, G.

« Rationality », dans Smith et Osherson [1995], Chapter 6, p. 175-211.

HITCHCOCK, C.

[1993] « A Generalized Probabilistic Theory of Causal Relevance », *Synthese*, 97, 3, p. 335-364.

[1996] « Causal Decision Theory and Decision-Theoretic Causation », *Noûs*, p. 508-526.

HOFSTADTER, D.

[1985] *Metamagical Themas : Questing for the Essence of Mind and Pattern*, Basic Books, New York.

HOLLAND, J. H., HOLYOAK, K. J., NISBETT, R. E. et THAGARD, P. R. (dirs.)

[1986] *Induction : Processes of Inference, Learning and Discovery,* Cambridge, MIT Press, Bradford books.

HOLLIS, M., et LUKES, S., (dirs.)

[1982] *Rationality and Relativism,* Oxford, Basil Blackwell ; 2e édition, Cambridge, MIT Press, 1991.

HOOKER, C. A., LEACH, J. J et MC LENNEN E. F., (dirs.)

[1978] *Foundations and Applications of Decision Theory*, Dordrecht, Reidel.

HORGAN, T.

[1981] « Counterfactuals and Newcomb's problem », *Journal of Philosophy*, vol. 78, p. 331-356.

HOWSON, C. et URBACH, P.

[1989] *Scientific Reasoning: The Bayesian Approach,* Chicago, Open Court.

JEFFREY, R. C.

[1965] *The Logic of Decision,* Chicago, University of Chicago Press ; la 1ère édition date de 1965 ; nous citons toujours la 2e édition, substantiellement révisée, qui date de 1983, avec les corrections de la réimpression de 1990.

[1965b] « New Foundations for Bayesian Decision Theory », d'abord publié dans Y. Bar-Hillel, *Logic, Methodology, Philosophy of Science*, Elsevier Science ; nous citons le texte reproduit comme chapitre 13 de Jeffrey [1992], p. 213-225.

[1968] « Probable Knowledge », chapitre 3 de Jeffrey [1992], p. 30-43

[1973] « Carnap's Inductive Logic », *Synthese*, 25, 299-306.

[1974] « Frameworks for Preference », chap. 14 de Jeffrey.[1992].

[1974b] « Preferences among Preferences », chap, 8 de Jeffrey [1992].

[1978] « Axiomatizing the Logic of Decision » chap. 15 de Jeffrey [1992].

[1981] « The Logic of Decision Defended », *Synthese*, vol. 48, #2, p. 473-492.

[1983] «Bayesianism with a Human Face», dans J. Earman, *Midwest studies in the philosophy of science*, vol. X, p. 133-156. Minneapolis, University of Minnesota Press ; reproduit dans Jeffrey [1992], p.77-107.

[1988] « How To Probabilize a Newcomb Problem », dans Fetzer [1988], p. 241-251.

[1988]a « Conditioning, Kinematics, and exchangeability », chap 7 de Jeffrey [1992], p. 108-117.

[1991] Postcript to « Probable Knowledge », voir Jeffrey [1992], p. 29 ; il s'agit d'un ajout dans cette réédition à un article datant de 1968.

[1992] *Probability and the Art of Judgment*, Cambridge, Cambridge University Press.

[1992a] « Probability and the Art of Judgment », article publié d'abord dans Chinstein et Hannaway [1985], nous citons le texte reproduit comme chapitre 4 de Jeffrey [1992] qui porte le même titre.

[1992b] « New foundations for Bayesian decision theory », chap 13 de Jeffrey [1992], p. 213- 225 ; version révisée d'un article paru en 1965.

[1993] « Causality in the Logic of Decision », *Philosophical Topics*, University of Kansas Press, vol. 21, no. 1, p. 139-151

[1995] « Probability Reparation : The Problem of New Explanation », *Philosophical Studies,* 77, p. 97-101.

[1996] « Decision Kinematics » Chap. 1 de Arrow et al. [1996], p. 3-19.

[1997] « Probabilistic Judgment » ; document disponible sur le site web de l'auteur, Princeton University ; version du 25/02/97 marquée : « version corrigée, augmentée et incomplète de *Probabilistic Thinking* (1995) ». http://www.princeton.edu/~bayesway.

[1997]b « Unknown Probabilities : In memory of Annemarie Anrod Shimony(1928-1995) », *Erkenntnis* 45 : 327-335.

[1999] « I was a Teenage Logical Positivist (Now a Septuagenarian Radical Probabilist », PSA Presidential Address, Kansas City, « in memory of Peter (C. G.) Hempel (1905-1997) » document disponible sur le site web de l'auteur, Princeton University ; version révisée, 04/01/99. http://www.princeton.edu/~bayesway

[2002] « Epistemology Probabilized » ; version pdf d'un texte à paraître dans B. Bouchon-Meunier, J. Guitterez-Rios, L. Magdalena, and R.R. Yaeger, dirs., *Technologies for Constructing Intelligent Systems*, Springer.

[2004] *Subjective Probability*, Cambridge, Cambridge University Press.

JEFFREYS, H.

[1939] *Theory of Probability*, Oxford; 3^{e} édition, 1961.

JONIDES, J.

[1995] « Working Memory and Thinking », chap. 7 de E. E. Smith et D. N. Osherson, [1995], p. 215-265.

JOYCE, J.M.

[1999] *The Foundations of Causal Decision Theory*, Cambridge, Cambridge Studies in Probablility, Induction, and Decision Theory, Cambridge University Press.

[2002] « Levi on Causal Decision Theory and the Possibility of Predicting One's Own Actions » *Philosophical Studies*, 110 , 69-102.

[2003] «Bayes's Theorem», *Stanford Encyclopedia of philosophy*, http://www.plato.standford.edu.

[2004] « Probabilistic Belief Revision and The Law of Conditional Excluded Middle », notes for a talk, n.p. University of Western Ontario : January, 10, 2004.

KANT, E.

[1781] *Critique de la Raison Pure*, (2^{e} édition, 1787) ; traduction de A. Tremesaygues et B. Pacaud, Paris, PUF, 1944.

KAPLAN, M.

[1996] *Decision Theory as Philosophy*, (édition révisée, 1998) ; Cambridge University Press.

KEENEY, R.L.

[1992] *Value-Focused Thinking : A Path to Creative Decisionmaking*, Cambridge, Harvard University Press.

KAHNEMAN, D., SLOVIC, P., TVERSKY, A., (dirs.)

[1982] *Judgement Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press, réimprimé en 1991.

KEYNES, J.M.

[1921] *A Treatise on Probability*, Macmillan & Co., London and New York; 2^{e} éd. 1929.

KONGUETSOF, L.

[1969] *Calcul différentiel et intégral*, McGraw-Hill, Montréal .

KYBURG, H. E.

[1968] « Bets and Beliefs », *American Philosophical Quaterly*, p. 63-78; nous citons le texte reproduit dans Gardenfors, P. et Sahlin, N. [1988], p. 101-117.

[1987] « Objective Probabilities », *IJCAI*, Vol. 2, p. 924-931.

[1991] « Normative and Descriptive Ideals », chap. 6 de Cummins and Pollock [1991], p. 129-140.

KYBURG, H., et SMOKLER, H., (dirs.)

[1964] *Studies in Subjective Probability*, New York, John Wiley.

LAPLACE, P. S. (de)

[1774] « Mémoire sur la probabilité des causes par les événements », *Mémoires présentés à l'Académie des Sciences*, VI, p. 621-656.

[1814] *Essai philosophique sur les probabilités*, Paris, Courcier, 5^{e} édition (1825) ; ce texte est devenu l'introduction de *Théorie Analytique des probabilités.*

[1878-1912] *Œuvres Complètes*, Paris, 14 volumes.

LEDWIG, M.

[1998] « The rationality of Probabilities for Actions in Decision Theory », *20^{e} Congrès Mondial de Philosophie*, Boston Mass. dans Paideia, « Philosophy of Action » :www.bu.edu/wcp/

LEHRER, K. et McGEE, V.

[1991] « An Epistemic Principle which Solves Newcomb's Paradox », *Grazer Philosophische Studien*, 40 ; avec 3 addenda, p. 197-232.

LEPAGE, F.

[1997] « Imaging, Conditional and Subjective Probability », *Dialogue*, vol. XXXVI, p. 113-135.

[2000] « A Many-Valued Probabilistic Conditional Logic », *Poznan Studies in the Philosophy of the Sciences and the Humanities*, vol. 71, p. 36-48.

LEPAGE, F., PAQUETTE, M. et RIVENC, F. (dirs.)

[2002] *Carnap Aujourd'hui*, Collection Analytiques, no. 14. Bellarmin Vrin.

LEVI, I.

[1974] « On Indeterminate Probabilities » *The Journal of Philosophy*, LXXI, 13, p. 391-418 ; nous citons le texte reproduit comme Chap.5 de Levy [1997].

[1983] « Review of Carnap and Jeffrey [1971] », *Philosophical Review*, 92, p. 20-21.

[1986] *Hard Choices,* Cambridge, Cambridge University Press.

[1997] *The Covenant of Reason: Rationality and the Commitments of Thought*, Cambridge, Cambridge University Press.

[2000] « Review of Joyce [1999] », *Journal of Philosophy*, 97, 7, p. 387-402.

LEWIS, C. I.

[1947] *Analysis of Knowledge and Valuation*, LaSalle, Illinois, Open Court.

LEWIS, D. K.

[1973] *Counterfactuals*, Library of Philosophy and Logic, Oxford, Basil Blackwell.

[1976] « Probabilities of Conditionals and Conditional Probabilities », Philosophical Review, 85, p. 297-315 ; aussi dans Lewis, D. [1986], p. 133-152.

[1979] « Prisoner's dilemma is a Newcomb Problem », *Philosophy and Public Affairs*, 8, p. 235-240 ; aussi dans Lewis [1986], p. 299-304.

[1979]a « Attitudes *De Dicto* and *De Se* » *The Philosophical Review*, 88, p. 513-543 ; aussi publié avec un *postcriptum* dans Lewis [1983], p. 133-160.

[1980] « A Subjectivist's Guide to Objective Chance », dans R.C. Jeffrey, éd. *Studies in Inductive Logic and Probability*, Vol. II, University of California Press ; aussi publié avec un *postcriptum* dans Lewis [1986], chapitre 19, p. 83 - 132.

[1981] « Causal Decision Theory », *Australasian Journal of Philosophy*, 59, 1981, p. 5-30, réimprimé avec un post-scriptum dans Lewis [1986], p. 305-339 ; aussi dans Moser [1990], chap 10, p. 235-263.

[1983] *Philosophical Papers*, Volume I, New York, Oxford University Press.

[1986] *Philosophical Papers*, Volume II, New York, Oxford University Press.

[1986b] *On The Plurality of Worlds*, Oxford, Basil Blackwell.

LINDSTRÖM, S. et RABINOWICZ, W.

[1989] « On Probabilistic Representation of Non-probabilistic Belief Revision », *Journal of Philosophical Logic*, 18, p. 69-101.

[1992] « Belief revision, epistemic conditionals and the Ramsey test », *Synthese*, 91, p. 195-237.

[1999] « Belief change for introspective agents », texte pdf; dedicated to P. Gärdenfors on his 50 th birthday.

LOAR, B.

[1980] « Ramsey's theory of belief and truth », dans Mellor [1980], p. 49-70.

LOUI, R. P.

[1993] « How a formal theory of rationality can be normative », Volume XC, no. 3, p. 137-143.

LUCE, R. D. et KRANTZ, D.

[1971]« Conditional Expected Utility », *Econometrica*, 39, p. 253-271.

LUCE, R. D. et RAIFFA, H.

[1957] *Games and Decisions, Introduction and Critical Survey*, Mineola, Dover Publications, 1989. (première édition, New York, Wiley, 1957).

MAHER, P.

[1993] *Betting on Theories*, Cambridge, Cambridge University Press.

MAITZEN, S. et WILSON, G.

[2003] « Newcomb's Hidden Regress », *Theory and Decision*, 54, No. 2, p. 151-162.

MARGALIT, A. et YAARI, M.

[1996] « Rationality and Comprehension », chap. 4 de Arrow et al. [1996], p. 89-100.

MAY, K. O.

[1954] «Intransitive utility and the aggregation of preference patterns», *Econometrica,* 22, p. 1-13.

McCALL, S.

[1987] « Decision », *Canadian Journal of Philosophy*, Volume 17, No. 2, juin, p. 261-288.

[1990] « Choice Trees », in Dunn and Gupta [1990] p. 231-244

[1994] *A Model of the Universe : Space-Time, Probability, and Decision,* Clarendon library of logic and philosophy, Oxford, Clarendon Paperbacks.

[1999] « Deliberation Reasons and Explanation Reasons », in R. Jackendoff et al. (dirs) *Language, Logic, and Concepts,* Bradford Books, MIT, pp. 97-108.

McCLENNEN, E. F.

[1990] *Rationality and Dynamic Choice: Foundational explorations* Cambridge, Cambridge University Press.

MCDERMOTT, D. A.

[1987] « Critique of pure reason ». *Computational Intelligence*, 3, p. 151-160. reproduit dans Boden [1990], p. 206-230.

McGEE, V.

[1991] « We Turing Machines aren't expected-utility maximizers (even ideally) », *Philosophical Studies*, 64, p. 115-223.

[1996] « Learning the impossible », dans Eells et Skyrms [1996].

[1999] « An Airtight Dutch Book », *Analysis*, 59, p. 257-265.

MELLOR, D. H.

[1980] *Prospects for Pragmatism : Essays in Honor of F.P. Ramsey*, Cambridge, Cambridge University Press.

[1990] « F. P. Ramsey », *Philosophy*, 70, p. 243-262 ; texte de l'émission de la BBC, « Better than the stars » consacrée Ramsey, en date du 27/02/1978. : http://www.dar.cam.ac.uk/~dhm11/RamseyLect.html

MILLER, G. A.

[1956] « The magical number seven, plus or minus two : Some limits on our capacity for processing information », Psychological Review, 63, p. 81-97.

MILLGRAM, E. et THAGARD, P.

[1996] « Deliberative Coherence », *Synthese*, 108, p. 63-88

MONTAGUE, R.

[1974] *Formal Philosophy, selected papers of Richard Montague*, Thomason, R., éd., New Haven, Yale University Press.

[1969] « On the Nature of Certain Philosophical Entities », *The Monist,* 53, p. 159-94 (1969) reproduit dans Thomason, R., éd., *Formal Philosophy, selected papers of Richard Montague* , p. 148-187, New Haven, Yale University Press, 1974.

MORGAN, C.

[2000] « Canonical Models and Probabilistic Semantics », *Poznan Studies in the Philosophy of the Sciences and the Humanities*, vol. 71, p. 17-35.

[2002] «Les probabilités comparatives comme fondement de la logique », dans Lepage et al. [2002], p. 177-202.

MOSER, P. K. (dir.)

[1990] *Rationality in Action: Contemporary approaches*, Cambridge, Cambridge University Press.

MUNDY, B.

[1989]« Elementary Categorial Logic, Predicates of Variable Degree, and Theory of Quantity », *Journal of Philosophical Logic*, 18, p. 115-140.

MUNIER, B. R.

[1988] dir., *Risk, Decision and Rationality*, Dordrecht, D.Reidel.

NAGEL, E., SUPPES, P. et TARSKI, A., (dirs.)

[1962] *Logic, Methodology and Philosophy of Science*. Stanford, Stanford University Press.

NOZICK, R.

[1970] « Newcomb's problem and two principles of choice », dans Rescher, N. [1970], p. 114-46 ; reproduit dans Moser [1990].

[1974] « Reflections on Newcomb's Problem », guest column in Martin Gardner's « Mathematical Games » series in *Scientific American*, (voir Gardner [1973]) (March 1974), p,.102-104, 106, 108 ; reproduit dans Nozick [1997], chap. 3 p. 74-84.

[1993] *The Nature of Rationality*, Princeton, Princeton University Press.

[1997] *Socratic Puzzles*, Cambridge, Harvard University Press.

NUTE, D.

[1980] *Topics in Conditional Logic*, Philosophical Studies Series in Philosophy, Vol. 20, D. Reidel, Dordrecht.

OSHERSON, D.

[1995] « Probability Judgment », dans Smith et Osherson [1995], Chapter 2, p. 35-76.

PAQUETTE, M.

[2001] « Compte rendu de : Dupuy, J.-P. et Livet, P. [1997] », *Dialogue*, XL, 1, p. 205-209.

[2002] « L'explication du choix rationnel chez Carnap », dans Lepage et al. [2002], p. 59-86.

PARIS, J. B.

[1994] *The Uncertain Reasoner's Companion : A Mathematical Perspective*, Cambridge Tracts in Theoretical Computer science, 39, Cambridge, University Press.

PEARL, J.

[1988] *Probabilistic Reasoning in Intelligent Systems*, San Mateo, Morgan Kaufmann.

[2000] *Causality, Models of Reasoning and Inference*, Cambridge University Press ; il existe des versions révisées de quelques chapitres sur le site web de l'auteur.

[2001] « Bayesianism and Causality, or, Why I am only a half-bayesian » dans Corfield et Williamson [2001], p. 19-36.

PETERSON, M.

[2004] « From Outcomes to Acts : A Non-Standard Axiomatization of Expected Uility », *Journal of Philosophical Logic*, August 2004, vol. 33, no. 4, p. 361-378.

PETTIT, P.

[1991] « Decision Theory and Folk Psychology », in Baccharach, M. and Hurley, S. [1991], p. 147-175.

PFANZAGL, J.

[1967] « Subjective Probability Derived from Morgenstern - Von Neumann Utility Concept », dans M. Shubik (dir.) *Essays in Mathematical Economics : In Honor of Oskar Morgenstern*, p. 237-251, Princeton, Princeton University Press.

PICAVET, E.

[1996] *Choix Rationnel et Vie Publique*, coll. Fondements de la Politique, Paris, PUF.

POLLARD, D.

[1999] « A modicum of measure theory », Notes du cours Statistics 330/600(Probability), Yale.

[2001] *User's Guide to Measure Theoretic Probability*, Cambridge University Press.

POLLOCK, J.

[1991] « OSCAR : A general Theory of Rationality », chapitre 9 de Cummins et Pollock [1991], p. 189-214.

POUNDSTONE, W.

[1992] *Prisoner's Dilemma*, New York, Doubleday.

PRIOR, A.

[1956] « The Consequences of Actions », *Proceedings of the Aristotelian Society, Supplementary Volume 30.*

[1967] *Past, Present and Future*, Oxford University Press,

RABINOWICZ, W.

[2002] « Does Practical Deliberation Crowd Out Self-Prediction », *Erkenntnis*, vol. 57, p. 91-122.

RAMSEY, F. P.

[1921] « The nature of propositions », avec la mention « Paper read to the Moral Sciences Club in Cambridge on 18. nov. 1921 » , reproduit dans Rescher [1991], p. 107-119.

[1922] « Mr. Keynes on Probability », *The Cambridge Magazine*, 11, p. 3-5, reproduit dans *The British Journal for the Philosophy of Science*, 40, (1989), p. 219-222.

[1922]a « On the Hypothesis of Limited Variety », dans Ramsey [1991] p. 269-272.

[1922]b « Induction : Keynes and Wittgenstein », dans Ramsey [1991], p. 296-301.

[1925] « Universals », dans Ramsey [1978], p. 17-39.

[1926] « Truth and Probability », texte rédigé en 1926 et publié pour la première fois en 1931, nous citons le texte d'après le chapitre 3, p. 58-100 de Ramsey [1978]. Il est réédité avec des modifications dans Ramsey [1990].

[1927] « Facts and propositions », dans Ramsey [1978], p. 40-57.

[1928] « Reasonable Degree of Belief » dans Ramsey [1990], p. 97.

[1928]b « Chance » dans Ramsey [1990], p. 104.

[1928]c « Law and Causality », chap. 6 de Ramsey [1978], p. 128-151.

[1929] « Probability and Partial Belief », dans Ramsey [1990], p. 95.

[1929]b « General Propositions and Causality », dans Ramsey [1978], p. 133-151.

[1929]c « Theories », dans Ramsey, F. P. [1978], p. 101-125.

[1978] *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, D. H. Mellor, éd., International Library of Psychology, Philosophy and Scientific Method, Humanities Press ; édition revue et augmentée d'un ouvrage édité par R. B. Braithwaite et paru en 1931 sous le titre: *The Foundations of Mathematics and other Logical Essays*.

[1990] *Philosophical Papers*, D. H. Mellor, éd., Cambridge, Cambridge University Press ; reprise revue et augmentée de Ramsey [1978].

[1991] *Notes on Philosophy, Probability and Mathematics*, recueil édité par M.C. Galavotti, Naples, Bibliopolis.

RESCHER, N., (dir.)

[1968] *Studies in Logical Theory*, Oxford, Blackwell.

[1970] *Essays in Honor of C.G. Hempel*, Dordrecht, D. Reidel.

RESCHER, N., et MAJER, U. (éds.)

[1991] *On Truth : Original Manuscript Materials (1927-1929) from the Ramsey Collection at the University of Pittsburgh*, Dordrecht, Kluwer Academic Publishers, coll. Episteme, vol. 16.

RESNIK, M. D.

[1987] *Choices, An Introduction to Decision Theory,* Minneapolis and London, University of Minnesota Press.

RIVENC, F.

[2002] « Carnap sur l'explication et la formation des concepts », chapitre 1 de Lepage et al. [2002].

RUNES, D. D., (dir.)

[1962] *Dictionary of Philosophy*, Totowa, NJ, Littlefield, Adams & Co.

RUSSELL, B.

[1903] *Principles of Mathematics*, 2e édition, 1938, W.W. Norton & Company, New York.

[1921] *The Analysis of Mind*, London.

[1959] *My Philosophical Development*, New York, Simon and Schuster ; traduit en français par Georges Auclair et paru sous le titre *Histoire de mes idées philosophiques*, Paris, Gallimard, coll. Tel. (1961).

RUSPINI, E.

[1987]« Epistemic Logics, Probability and the Calculus of Evidence », *IJCAI*, Vol. 2, p. 924-931.

SAHLIN, N.-E.

[1990] *The Philosophy of F. P. Ramsey*, Cambridge, Cambridge University Press.

SALMON, W. C.

[1970] « Statistical Explanation », dans Colodny, *The Nature and Function of Scientific Theories*, Pittsburgh University Press, Pittsburgh, 1970 ; reproduit dans Salmon, W. C., Greeno, J. G. et Jeffrey, R. C., *Statistical Explanation and Statistical Relevance*, Pittsburgh University Press, Pittsburgh,1971, 29-87.

[1988] « Dynamic Rationality: Propensity, Probability and Credence », dans Fetzer [1988], p. 3-40.

SAVAGE, L.

[1954] *The Foundations of Statistics,* John Wiley & Sons; nous citons la seconde édition révisée et augmentée, Dover, 1972.

[1967] « Difficulties in the Theory of Personal Probability », *Philosophy of Science*, 34, p. 306-310.

[1971] « Elicitation of personal probabilities and expectations », *Journal of the American Statistical Association*, 65, p. 1501-1598.

SCHICK, F.

[1997] *Making Choices : A Reconstruction of Decision Theory*, Cambridge, Cambridge University Press.

SCHILPP P. A.

[1963] *The Philosophy of Rudolf Carnap*, LaSalle, Ill., Open Court.

SCHLEE, E.E.

[1997] «The Sure Thing Principle and the Value of Information», *Theory and Decision*, Vol. 42, p. 21-36.

SCHORER, P.

[2003] « Simulation Paradoxes », texte pdf, Hewlett Packard Laboratories, Palo Alto.

SHAFER, G.

[1986] « Savage revisited », *Statistical Science*, 1, p. 463-485 ; reproduit dans Bell, D. E., Raiffa, H. and Tversky, A. [1988], page 193-234.

SHENOY, P.

[1988]« Game trees for Decision Analysis », *Theory and Decision*, 44, p. 149-171.

SEARLE, J.

[2001] *Rationality in Action : The Jean Nicod Lectures*, Bradford Books, MIT.

SHIMONY, A.

[1967] « Amplifying Personal Probability Theory : Comments on L. J. Savage [1967] », *Philosophy of Science*, 34, p. 326-332.

[1955] « Coherence and the Axioms of Confirmation », *Journal of Symbolic Logic*, Vol. 20.

SIMON, H. A.

[1955] « A Behavioral Model of Rational Choice », *Quaterly Journal of Economics*, vol.69.

[1981] *The Sciences of the Artificial*, 2e édition, Cambridge.

[1983] *Reason in Human Affairs*, Stanford University Press.

SKYRMS, B.

[1980] *Causal Necessity : A Pragmatic Investigation of the Necessity of Laws*, New Haven and London, Yale University Press.

[1982] « Causal decision theory », *Journal of Philosophy*, Vol. 79, p. 695-711.

[1984] *Pragmatism and Empiricism*, New Haven and London, Yale University Press.

[1986] *Choice & Chance: An introduction to Inductive Logic*, 3e édition, Belmont, Wadsworth.

[1986]a « Deliberational Equilibria », *Topoi*, 5, p. 59-66.

[1987] « Dynamic Coherence and Probability Kinematics », *Philosophy of Science*, 54, p. 1-20.

[1988] « Deliberational Dynamics and the Foundations of Bayesian Game Theory », *Philosophical Perspectives*, Tomberlin, J. E, (dir.), 2, « Epistemology », p. 346-367.

[1990] *The Dynamics of Rational Deliberation*, Cambridge, Harvard University Press.

[1990]a « Ratifiability and the Logic of Decision », dans French, P. A., Uehling, T. E., Wettstein, H. K., [1990], p. 44-57.

[1996] « Commitment », chap. 2 de *Evolution of the Social Contract*, Cambridge University Press, p. 22-44.

[1997] « The Structure of Radical Probabilism », *Erkenntnis*, 45, p. 285-297.

SLOTE, M.

[1989] *Beyond Optimizing: A Study of Rational Choice,* London, Harvard University Press.

SMITH, E., LANGSLOR, C. et NISBETT, R.

[1992] « The Case for Rules in Reasonning », *Cognitive Science*, vol. 16, no. 1, p. 1-41.

SMITH, E. E. et OSHERSON, D. N. (dirs.)

[1995] *Thinking : An Invitation to Cognitive Science*, Vol. 3, 2e éd., Cambridge, Mass., Bradford Books, MIT Press.

SOBEL, J. H.

[1986] « Notes on decision theory : Old wine in new bottles » *Australasian Journal of Philosophy*, 64 p. 407-437; aussi dans Sobel [1994], chapitre 8, p. 141-173.

[1988] « Infallible Predictors » *Philosophical Review*, 97, p. 3-24 ; aussi dans Sobel [1994], chapitre 5, p. 100-118.

[1989] « Kent Bach on good arguments » *Canadian Journal of Philosophy*, 20, p. 447-453 ; aussi dans Sobel [1994], chapitre 6, p. 119-125.

[1990] « Newcomblike Problems » *Midwest Studies in Philosophy*, 15, p. 224-255.

[1994] *Taking Chances: Essays on Rational Choice,* Cambridge, Cambridge Studies in Probablility, Induction, and Decision Theory, Cambridge University Press.

[1998] « Ramsey's Foundations Extended to Probabilities », *Theory and decision*, 44, p. 231-278.

SPOHN, W.

[1977] « Where Luce and Krantz Do Really Generalize Savage's Decision Model », *Erkenntnis*, 11, p. 113-134.

STALNAKER, R. C.

[1968] « A Theory of Conditionals », dans Rescher [1968].

[1984] *Inquiry*, Cambridge, Mass. et Londres, Bradford books, MIT Press,.

[1999] *Context and Content : Essays on speech and thought*, Oxford Cognitive Science Series, Oxford Unniversity Press.

SUPPES, P.

[1972] *Axiomatic Set Theory*, Dover Publications, New York ; version revue et augmentée de l'édition originale, (1960), Van Nostrand.

[1981] *Logique du probable : Démarche bayésienne et rationalité*, Nouvelle Bibliothèque Scientifique, Paris, Flammarion.

[1984] *Probabilistic Metaphysics*, Oxford, Basil Blackwell.

SUPPES, P. et DAVIDSON, D.

[1956] « A Finistic Axiomatisation of Subjective Probability and Utility », *Econometrica*, Vol. 26, 1956, p. 264-275.

SWINBURNE, R.

[1974] (éd.) *The Justification of Induction*, Oxford Readings in Philosophy, Oxford University Press.

[1986] « The Indeterminism of Human Actions », dans French, P. A. Uehling, T. E. et Wettstein, H. K. *Midwest Studies in Philosophy, X*, p. 431-449.

TALBOTT, W.

[2001] «Bayesian Epistemology», *Stanford Encyclopedia of philosophy*, http://www.plato.standford.edu.

TARSKI, A.

[1956] *Logic, Semantics, Meta-mathematics*, trad. angl. J. H. Woodger, 2e édition de J. Corcoran, Indianapolis, Hackett.

THOMASON R. H.

[1970] « Indeterminist time and truth-value gaps », *Theoria*, 3, p. 264-281.

[1984] « Combinations of Tense and Modality », dans Gabbay et Guenthner, dirs., *Handbook of Philosophical Logic, Vol. II : Extensions of Classical Logic*, D. Reidel, p. 135-165.

THOMASON R. H. et GUPTA, A.

[1980] « A Theory of Conditional in the Context of Branching Time », *Philosophical Review*, 89, p. 65-90.

THOMASON R. H. et HORTY, J. F.

[1996] « Nondeterministic Action and dominance : Foundations for planning and qualitative decision » dans Shoham, Y. (dir.) *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge*, TARK-96, Morgan-Kaufmann p. 229-250 ; il existe une version corrigée, (web), n.p., document pdf, 24 p.

TODHUNTER, I.

[1865] *A History of the Mathematical Theory of Probability from the time of Pascal to that of Laplace* ; New York, repr. Chelsea Press, 1949.

TVERSKY, A.

[1969] «Intransitivity of preferences», *Psychological Review*, 76, p. 31-48.

VAN FRAASSEN, BAS C.

[1980] *The Scientific Image*, Oxford, Clarendon Press.

[1986] « A Demonstration of the Jeffrey Conditionalization Rule » *Erkenntnis*, 24, p.17-24.

[1989] *Laws and Symmetry,* Oxford, Oxford University Press.

[1992] « Faire figure dans un monde probabiliste », dans D. Laurier et F. Lepage, (dirs.) *Essais sur le langage et l'intentionalité*, Montréal et Paris, Bellarmin-Vrin, coll. Analytiques, #4, 1992, p. 307-321 ; Une version différente est parue dans Dunn et Gupta [1990].

VANDERVEKEN, D.

[1990] *Meaning and Speech acts*, Vol I. *Principles of Language Use* and Vol. II *Formal Semantics of Succes and Satisfaction.* Cambridge, Cambridge University Press.

[1995] « A New Formulation of the Logic of Propositions », dans M. Marion et R.S. Cohen (éds.), *Québec Studies in the Philosophy of Science* I, p. 95-105.

[2001] « *Foundations of the Logic of Actions* », texte photocopié, version augmentée de « The Basic Logic of Action », *Cahiers d'épistémologie*, Université du Québec à Montréal, Cahier no. 9907, 1999.

[2001]a « Illocutionary logic and discourse typology », *Revue Internationale de Philosophie*, 2, no. 217, p. 243-255.

[2002] « Towards the Foundations of the Logic of Action », UQTR, n.p. 49 pages.

VELLEMAN, J. D.

[1993] « The Story of Rational Action », *Philosophical Topics*, vol. 21, no. 1

[2000] *The Possibility of Practical Reason*, Oxford University Press.

VICKERS, J. M.

[1988] *Chance and Structure : An Essay in the Logical Foundations of Probability.* Oxford : Clarendon Press.

[1993] « Probability and Utility », Chap. 6 de Dubucs [1993], p. 109-127.

[2001] « Logic, Probability, and Coherence », *Philosophy of Science*, 68, p. 95-110.

VON NEUMANN, J.

[1963] *Method in the Physical Science*, Collected Works, Vol. 6, Pergamon Books.

VON NEUMANN, J. et MORGENSTERN O.

[1944] *Theory of Games and Economic Behavior*, Princeton, Princeton University Press ; 3^{e} édition, 1953.

VON PLATO, J.

[1994] *Creating Modern Probability : Mathematics, Physics and Philosophy in Historical Perspective.* Cambrige Studies in Probability, Induction, and Decision Theory, Cambridge University Press.

WAKKER, P.

[1993] « Savage's Axioms Usually Imply Violation of Strict Stochastic Dominance », *Review of Economic Studies*, Vol. 60, no. 2, p. 487-493.

WALLISER, B. et ZWIRN, D.

[1997] « Les Règles de Révision des Croyances », dans Dupuy, J.-P. et Livet, P. [1997], p. 190-222.

WEIRICH. P.

[2000] « Review of James Joyce, *The Foundations of Causal Decision Theory* », *Philosophical Books*, Vol. 41, p. 217–219.

[2001] *Decision Space : Multidimensional Utility Analysis*, Cambridge Studies in Probability, Induction, and Decision Theory, Cambridge University Press.

WHITEHEAD, A. N. et RUSSELL, B.

[1910] *Principia Mathematica*, (1962) paperback edition to *56, Cambridge et New York, Cambridge University Press.

YEGHIAYAN, E.

[2000] « Richard Jeffrey : A Bibliography compiled by Eddie Yeghiayan », n.p., University of California, 34 p.

YUDKOWSKY, E. S.

[2002] « Levels of Organization in General Intelligence », à paraître dans Goertzel, B. et Pennachin, C., *Real AI: New Approaches to Artificial General Intelligence.*

ZABELL, S. L.

[1997] « Confirming Universal Generalizations », *Erkenntniss*, 45, p. 267-283.

NOTES ET RÉFÉRENCES

[1] von Neumann, [1963].

[2] Davidson [1976] p. 176.

[3] Skyrms [1996], p. 41.

[4] Dans un passage *Reason and Human Affairs* qui est reproduit dans Moser [1990], chap. 8, p. 189.

[5] Cette méthodologie a été décrite en détail par R. Carnap dans Carnap [1950-4] et par Alfred Tarski dans plusieurs articles. Voir, par exemple, Tarski [1956], chap. VIII, XV et XVI. Voir Rivenc [2002] chap. 1 de Lepage et al. [2002].

[6] On peut trouver des définitions comparables dans les dictionnaires d'économie. Pour le lecteur non-initié, nous recommandons la lecture de Hargreaves Heap [1992], chap. 1, Le recueil de Paul K. Moser [1990], ou, en français, la première partie de Picavet [1996].

[7] Arnauld et Nicole [1680], p. 428.

[8] C'est ainsi que nous désignerons les « décideurs » tout au long de la thèse.

[9] Voir Vanderveken [2001]a.

[10] von Neumann et Morgenstern [1944].

[11] Davidson [1974], discute de la difficulté d'interpréter les résultats empiriques de ses études qui décrivent des comportements de choix. Pour les préférences apparemment inconsistantes, par exemple, il suffit, de laisser tomber la contrainte de cohérence à travers le temps. Voir Loui [1993], qui tire de cet argument une conclusion sceptique.

[12] Henry E. Kyburg, dans Kyburg [1991] soutient un point de vue opposé. Selon lui la logique de la décision et la logique déductive sont toutes les deux des théories partiellement normatives et partiellement descriptives. Il considère que la situation est plus confuse en logique de la décision qu'en logique déductive.

[13] Lewis [1981], note 8.

[14] Nous mentionnons, à la volée, quelques titres qui figurent dans notre bibliographie : Tversky [1969] ; Allais [1953] et Allais [1979] ; Bell, D. Raiffa et Tversky [1988] ; Gigerenzer et Selten [2001]

[15] Lewis [1981], § 2, p. 237.

[16] McGee [1991].

[17] Cette position était celle de Savage. Voir Savage [1954] et Savage [1971].

[18] Searle [2001], p. 6.

[19] Davidson signale le caractère apparemment infalsifiable de la théorie dans Davidson [1974], p. 273.

[20] En réalité, Searle a donné une nouvelle formulation à une vieille énigme, à peine plus jeune que les premières axiomatisations du choix rationnel. Nous ne pouvons réclamer le crédit pour la solution car L. Savage et J. M. Joyce ont résolu ce type de difficulté antérieurement. Voir Savage [1954], p. 81 et Joyce [1999], p. 94.

[21] Joyce [1999], p. 72.

[22] Nous reprenons la formule que David Lewis utilise pour caractériser une position de E. Eells à laquelle Lewis ne souscrit pas. Voir Lewis [1981], § 4.

[23] Fishburn [1981].

[24] Comme nous le verrons, même les théorèmes de représentation qui sont invoqués pour valider la logique de la décision demandent à être interprétés.

[25] La seule exception est Leonard Savage qui était mathématicien. Dans son cas, il faut ajouter que la plupart des mathématiciens considéraient qu'il était un philosophe, nous aussi d'ailleurs, mais pour des raisons opposées.

[26] Issac Levi signale en de nombreux endroits que les théories causales de la décision ainsi que la théorie de Jeffrey sont surtout discutées en philosophie et peu ailleurs. Elle ne sont pas discutées, par exemple, en économie et en sciences de la décision.

Notes du chapitre 2 :

[1] Kant [1787], (1781, pour l'édition originale). Voir la Théorie transcendantale de la méthode, troisième section, « De l'opinion, de la science et de la foi », p. 554 de la traduction de A. Tremesaygues et B. Pacaud. Rendue sommairement, l'idée de Kant est que dans la *foi pragmatique*, on peut mesurer l'intensité de la conviction par la disposition à parier, par exemple, « un ducat » ou « toute une vie ». Pour une critique de cette méthode, voir Ramsey [1926], p. 74. Selon Picavet [1996], p. 172 *sq.*, « certaines des

thèses couramment attribuées à Ramsey sont en fait [...] dues à Kant ». Soit, mais nous pensons que l'extension de ce jugement doit s'arrêter à la porte du concept de probabilité. Suivant Hacking [1975], p. 14 et le consensus des autres sources, nous attribuons la paternité du concept contemporain de probabilité subjective (personnelle) à Ramsey et de Finetti. Seule la référence à Borel [1924], soulignée par Galavotti [1991] et [2005], introduit une nuance (légère) à ce jugement d'antériorité historique. Des indications plus complètes sont fournies plus bas, à la note 3 de la même page.

[2] Voir Borel [1924], p. 46-60 de Kyburg et Smokler [1964].

[3] C'est du moins l'avis de M. C. Galavotti ; voir Galavotti [1991], p. 239. Yvon Gauthier a attiré notre attention sur le fait que l'interprétation épistémique des probabilités était déjà présente chez Pierre Simon de Laplace. Voir son *Traité analytique des probabilités* (1812) repris dans *L'Essai philosophique sur les probabilités* (1814). Pour réconcilier l'avis de M. C. Galavotti et l'observation de Y. Gauthier, il faut peut-être souligner la différence entre « *interpréter* les probabilités de manière subjective » et « *définir* les probabilités comme des degrés de croyance ». Laplace distingue deux interprétations des probabilités qui correspondent respectivement à *l'espérance mathématique* de *l'espérance morale* pour résoudre le paradoxe de Saint-Pétersbourg. La conception des probabilités de Laplace est éclectique selon Galavotti et elle attribue la paternité de l'interprétation subjective à William F. Donkin (1814-1869). Voir Todhunter [1865] et Galavotti [2005], p. 64, p. 66 et p. 189. À ce compte, notons que Borel partageait aussi *le point de vue français* au sujet de la double nature des probabilités ; voir Von Plato [1994], p. 36.

[4] Voir les indications historiques de R. Jeffrey dans Jeffrey [1992a], p. 66, note 16.

[5] Ce sont respectivement les termes de Skyrms [1986], p. 202 et de Sahlin [1990], p. 3.

[6] Voir à ce sujet Ramsey [1928].

[7] Ramsey [1926], p. 82.

[8] Sobel [1998] défend et prolonge l'approche fondationnelle de Ramsey dans le contexte de la théorie causale de la décision.

[9] Voir Hacking [1975], chap. 2. Voir aussi Jeffrey [1992], (1985), chap. 4 pour une introduction historique à l'interprétation subjective. Jeffrey définit une conception de la probabilité et du jugement probable pour laquelle il utilise les termes « probabilisme » et « probabilisme radical » et il associe Ramsey à cette conception.

[10] Nous utiliserons l'expression « conception fréquentielle » pour désigner cette conception des probabilités qui correspond à l'interprétation en termes de fréquences. Pour un exposé et une discussion critique de cette conception et de ses variantes, voir van Fraassen [1980], chap. 6, en particulier, aux pages 181-187.

[11] Cet exemple est choisi parce qu'il semble simple. C'est un trait accidentel que les deux résultats se voient assigner la même probabilité et les deux interprétations pourraient être utilisées dans une situation où les résultats ne sont pas équiprobables, comme ce serait le cas pour la probabilité qu'il pleuve aujourd'hui. On doit également adopter une attitude critique envers l'impression qu'on peut posséder une connaissance *a priori* de la probabilité du résultat. Enfin, on dit parfois de ce genre d'exemples relatifs à des jeux de hasard, comme ceux qui utilisent un jeu de cartes ou un dé à six faces, qu'ils correspondent bien à la conception classique des probabilités. Voir la discussion dans Freund [1973], p. 40.

[12] Freund [1973], p. 47.

[13] Russell présente la problématique de l'inférence non démonstrative ainsi que ses principales idées sur la question dans Russell [1959], chap. XVI, p. 190 *sq.*

[14] Keynes [1921], p. 217.

[15] ibid., p. 220.

[16] Skyrms [1986], p. 12.

[17] Ramsey [1926] p. 99.

[18] Hughes Leblanc a cherché à développer une sémantique probabiliste pour une logique qui s'éloigne le moins possible de la logique classique. Entre autres résultats, il a montré que la logique probabiliste n'est pas une logique infinitaire — un calcul du premier ordre infiniment multivalent. Charles Morgan a proposé une sémantique pour la logique probabiliste qui est fondée sur le concept de probabilité comparative.

[19] Nous utilisons l'expression « induction naturelle » par opposition à « induction formelle » pour désigner l'inférence qui procède d'un certain nombre d'énoncés d'observation vers une proposition universelle qui exprime une loi générale. L'étude des conditions qui affectent la validité d'une telle inférence constitue le problème traditionnel de l'induction. L'induction naturelle est, par sa nature, risquée. Par opposition, l'expression « induction formelle » désigne un mode de preuve utilisé principalement en logique et dans l'arithmétique formalisée et cette forme de raisonnement est formellement valide au sens de la théorie de la déduction. L'induction formelle est un schéma de preuve.

[20] Skyrms [1986], p. 21.

[21] Skyrms [1986], p. 21.

[22] En plus de l'essai que nous discutons, « Truth and Probability », qui est la référence principale, il existe d'autres textes de Ramsey datés de 1922 où Ramsey discute et critique la théorie de Keynes. Ce sont les textes indiqués sous Ramsey [1922], Ramsey [1922a] et Ramsey [1922]b dans la bibliographie.

[23] « Between two sets of propositions, therefore, there exists a relation, in virtue of which, if we know the first, we can attach to the latter some degree of rational belief ». Keynes [1921], p. 6. Pour Keynes, cette relation logique est indépendante de la connaissance qu'en ont les humains et de leurs croyances à son sujet. Plus loin, p. 281, il introduit explicitement le terme de probabilité subjective pour désigner cette probabilité qui est dépendante de la connaissance ou de l'ignorance et qui est relative à l'esprit d'un sujet.

[24] Ramsey [1926], p. 62.

[25] *ibid.* p. 64.

[26] Ramsey [1926], p. 65.

[27] Les expressions « base de connaissances » et « sphère de croyances » sont utilisées ici selon leur sens intuitif. L'expression « base de connaissances » vient du domaine de l'intelligence artificielle et elle traduit l'expression « knowledge base ». Elle est manifestement construite sur le modèle de l'expression « base de données » et il s'agit en quelque sorte d'une généralisation de ce concept. L'expression « sphère de croyances » exprime la même idée dans le contexte de l'épistémologie et des théories formelles de la connaissance. L'image de la sphère permet de parler en termes de noyau, de périphérie, du degré de « centralité » d'une croyance, et ainsi de suite. Pour une explication technique d'un concept similaire, voir Gärdenfors [1988], chap. 2.

[28] Ramsey [1929]b, p. 143.

[29] On peut le démontrer aisément. Voir Sobel [1990], p.239, ou Jeffrey [1965], p. 214.

[30] Voir Stalnaker [1968] et la critique décisive de Lewis [1976]. La conditionnelle de Stalnaker peut s'interpréter intuitivement de la façon suivante, « Ajouter A à votre sphère de croyances en effectuant les modifications minimales qui s'imposent et évaluez C sur cette base ». La conditionnelle de Ramsey correspond au cas trivial où aucune révision n'est nécessaire. Voir aussi Ramsey [1929]b, p. 143, note 1, Nute [1980], chap. 1, Stalnaker [1984], chap. 6 et Gärdenfors [1988], chap. 7.

[31] Nous renvoyons le lecteur intéressé à la théorie de la confirmation dans une optique bayésienne aux discussions récentes de John Earman, Earman [1992], ainsi qu'à l'ouvrage remarquable de Colin Howson et Peter Urbach, Howson et Urbach [1993]. Pour une approche également récente mais dans l'optique de la psychologie cognitive, on peut

consulter Holland, Holyoak, Nisbett et Thagard [1989]. L'approche probabiliste en philosophie des sciences est brillamment exposée et soutenue dans van Fraassen [1989].

[32] Voir le passage de Keynes cité par Ramsey, Ramsey [1926], p. 65-66.

[33] Ramsey [1926], p. 68.

[34] Outre le concept de croyance partielle que nous discutons ici, la probabilité, l'utilité, la préférence sont comparables et ces concepts font l'objet de critiques similaires. On trouvera les arguments d'une défense de la légitimité de tels concepts dans von Neumann et Morgenstern [1944], p. 16 à 21, dans Jeffrey [1965], dans Skyrms [1986], p. 206 et dans van Fraassen [1989], p. 153-154 et *passim.*

[35] Suivant l'usage, nous traduisons « *efficacy* » par efficacité. En regard du sens, il vaudrait peut-être mieux parler « d'effectivité » causale.

[36] Cette critique se trouve dans *The Analysis of Mind*, Russell [1921].

[37] Il y a des arguments plus forts que ceux de Russell qui plaident contre la conception pragmatique de la croyance et le béhaviorisme qui semble en être la philosophie achevée. Nous aurons l'occasion de revenir sur l'analyse de la croyance dans le contexte de la théorie de la décision. Voir Jeffrey [1965], chap. 4, section 4.7 et Joyce [1999], p. 21 et *passim.*

[38] Ramsey [1926], p. 75.

[39] Ramsey n'expose pas la définition qu'il considère appropriée pour le concept d'espérance mathématique. Le passage de la p. 76 que nous citons suggère qu'il s'agit du concept habituel et c'est lui que nous paraphrasons. L'utilisation de ce concept dans le contexte généralisé de la théorie du choix rationnel a été critiquée. Il faut noter au passage qu'il y a des concepts qui sont des variantes de l'espérance mathématique. Ainsi, on mentionne le concept « d'espérance morale » introduit par Daniel Bernoulli et utilisé pour solutionner le paradoxe de Saint-Petersbourg par Pierre Simon de Laplace. De même, Sobel [1998], p. 233 parle « d'espérance évidentielle » et « d'espérance causale ».

[40] Ramsey [1926], p. 76.

[41] Nous introduisons des définitions explicites pour la terminologie des paris à la section 8 du présent chapitre.

[42] «[...] the person may have a special eagerness or reluctance to bet » Ramsey [1926], p. 74 et plus loin «[...] but this is vitiated, as already explained, by love or hatred of excitement » ibid. p. 78.

[43] Pour un exposé de la théorie de la croyance de Ramsey et des rapports entre cette théorie et celles de Russell et Wittgenstein, voir Loar [1980].

[44] La paraphrase de « A croit que *p* » donnée quelques lignes plus haut est fondée sur l'idée qu'une assertion et une croyance ayant le même contenu sont vraies dans les mêmes conditions. Voir Vanderveken [1990] p. 58. La seconde paraphrase est réminiscente d'une analyse proposée par R. Barcan Marcus dans quelques articles sur la croyance (1981, 1983, 1990). Voir aussi Jeffrey [1965], p. 68-70.

[45] Ramsey [1926], p. 79, note 1.

[46] C'est la proposition 4.01 du *Tractatus Logico-Philosophicus.*

[47] L'analogie est utilisée dans Ramsey [1927], p. 47 : « Nevertheless, just as in the study of chess nothing is gained by discussing the atoms of which the chessmen are composed, so in the study of logic nothing is gained by entering into the ultimate analysis of names and the objects they signify ».

[48] Ramsey [1927], p. 53.

[49] Relativement à « ne pas croire » on mentionne le verbe *discroire*, un néologisme proposé par Denis Vernant construit sur le modèle de *disbelief.* Discroire que *p*, c'est « ne pas croire que *p* » dans un sens qui n'est pas équivalent à « croire que non-*p* ». « A discroit que *p* » est vraie si et seulement si ce n'est pas le cas que A croit que *p*. A croit que non-*p* implique que A discroit que *p* mais la réciproque n'est pas vraie si A n'a aucune attitude doxastique relativement à *p*.

[50] idem, p. 52.

[51] Ramsey [1927], p. 51 et 53.

[52] Pour décrire la théorie de Ramsey sans trop la dénaturer, nous utilisons informellement le concept de proposition dans un sens qui est devenu courant en logique philosophique où on identifie une proposition avec un ensemble de mondes possibles. Pour une explication courte et dense de cet usage courant, voir Lewis [1973], p. 46-47, en particulier la note. Les solutions raisonnables pour remédier aux défauts de ce concept seront examinées plus loin. En particulier nous signalons la théorie prédicative des propositions telle que résumée et utilisée dans Vanderveken [2001].

[53] Elle forme un pré-ordre complet . Voir Fishburn [1970], p. 30.

[54] Ramsey [1926], p. 78-79. Ces mondes possibles jouent ici le rôle que nous assignerons aux *quasi-histoires* dans le contexte de la logique temporelle modale.

[55] Voir Davidson, D., Suppes, P. et Siegel, S. [1957]. Dans Davidson [1974], on trouve un bilan et une discussion critique de ces expériences de psychologie dans le contexte d'une discussion de la théorie de Ramsey.

[56] Ramsey [1926], p. 79.

[57] Voir Jeffrey [1965], p. 55-57.

[58] Voir Sobel [1998], p. 236.

[59] Voir Sahlin [1990], p. 231, note 4. Sa remarque renvoie à Ramsey [1929]b. Notons au passage que nous n'y avons pas trouvé les raisons de croire que Ramsey remet en cause le concept de proposition atomique. Il critique cependant l'atomisme logique de Russell dans Ramsey [1925]. Voir Mundy [1989], p. 117. Mundy développe une théorie comportant des prédicats à degrés variables, donnant suite à une suggestion de Ramsey.

[60] L'expression « dutch book » a été introduite par Isaac Levi en 1965 dans une conférence à l'APA selon Kyburg [1968], p. 104. Elle désigne une série de paris qui semblent acceptables pour un parieur donné mais qui sont en réalité systématiquement défavorables. L'idée se trouvait chez Ramsey et de Finetti [1935]. On parle du théorème du « dutch book » pour désigner la preuve qu'il existe une telle série de paris pour quiconque rejette tel ou tel axiome. La référence à l'habileté des preneurs au livre (*bookies*) hollandais est d'ordre folklorique.

[61] Énoncé simplement, le principe de tolérance permet à quiconque d'utiliser le langage qui convient le mieux à son propos. Il est formulé pour la première fois dans Carnap [1937], p. 51 de la façon suivante : « *It is not our business to set up prohibitions, but to arrive at conventions* ».

[62] Voir, par exemple, Carnap [1963a], p. 18.

[63] Le statut des termes théoriques est discuté par Ramsey dans l'article intitulé « Theories », Ramsey [1929]c.

[64] Dans cette phrase et la précédente, nous utilisons la distinction fort ancienne entre l'adéquation formelle et l'adéquation matérielle qui appartient désormais à la méthodologie des sciences déductives, à la suite de Tarski [1956].

[65] Voir Sobel [1998], section 3, p. 240-241.

[66] Cette formulation est réminiscente de la paraphrase de Nuel Belnap pour l'opérateur *stit* de la logique de l'action fondée sur « faire en sorte que ». Voir Belnap [1994], passim.

[67] Ramsey [1926], p. 80.

[68] Voir les définitions comparables de Sahlin [1990], p.17 et Jeffrey [1965], chap. 3.

[69] L'expression entre parenthèses est l'abréviation de l'expression « α est le monde possible qui diffère au plus du monde actuel au moment du pari par le fait que l'agent gagne cinq dollars et β est le monde possible qui diffère au plus du monde actuel par le fait que l'agent perd cinq dollars ».

[70] Selon sa formule, « [...] but this will not seem unreasonable when it is seen that all our lives we are in a sense betting ». Ramsey [1926], p. 85.

[71] Il n'y a pas d'inconvénient à ajouter que la relation d'accessibilité dans ce contexte est la relation d'accessibilité de la logique modale S5.

[72] Nous disons « ici » pour souligner que dans un autre contexte théorique, nous pourrions souhaiter *définir* le terme de monde possible en prenant, par exemple, le terme de proposition comme terme primitif. Il n'y aurait pas d'incohérence dans cette façon de procéder car d'un point de vue logique et méthodologique, le mode de construction d'une théorie est une question de commodité.

[73] Cette remarque est nécessaire pour éviter de prêter flanc aux critiques que l'on peut adresser à ce que Lewis appelle *l'ersatzisme linguistique*.Voir Lewis [1986]b, p. 142. Pour une discussion supplémentaire de l'*ersatzisme* fondée sur une suggestion de Quine, voir Lewis [1973] p. 84 sq.

[74] Pour être inutilement plus précis, il faudrait ajouter que le concept de vérité de cette définition est relatif à un modèle et que le concept de consistance (syntaxique) est relatif à un système pour le calcul des propositions. Le terme « proposition caractéristique » est repris de Storrs McCall. Dans Sobel [1998], Sobel utilise l'expression « proposition-monde » (*world-proposition*). Voir aussi le concept voisin *d'ensemble maximal d'énoncés* dans Chellas [1980], p. 60, et le concept de *proposition maximale vraie* dans Vanderveken [1991], vol. II, p. 85.

[75] Cette explication est due à Sobel [1998], section 4, pour le quasi-monde (*near-world)* $\alpha^{[+p]}$. On trouve une explication équivalente dans Jeffrey [1965], p. 56 pour une expression dont la notation est $\alpha + p$.

[76] Gärdenfors et Sahlin [1988], p. 6-7.

[77] L'exemple est discuté dans McClennen [1990], p. 64.

[78] À propos de la transitivité des préférences, voir la courte et éclairante discussion de Resnik dans Resnik [1987], p. 23-24 à partir de l'exemple d'une série de tasses de café dont chacune est légèrement plus sucrée que la précédente, mais où l'accroissement du

taux de sucre est en deçà du seuil de la sensibilité gustative humaine. Dans ce cas, la capacité d'avoir des préférences transitives semble dépasser nos capacités biologiques de discrimination. Peut-on raisonnablement demander que la contrainte de rationalité exige, dans ce cas, d'avoir recours à l'analyse chimique ? Dans le cas d'un état de santé qui commande d'éviter le sucre, il me semble que la réponse est « oui ».

[79] Voir Jeffrey [1965], p. 225, le dernier paragraphe énonce clairement le refus des préférences intransitives et nous l'adoptons comme une protestation à lire à tue-tête.

[80] La distinction essentielle entre le domaine de ce qui *raisonnable* et le domaine de ce qui est *rationnel* est proposée par Kyburg dans Kyburg [1968]. Il ne l'utilise pas pour discuter de la transitivité mais il est clair, comme il l'affirme, qu'elle est à l'œuvre chez Ramsey.

[81] Voir Whitehead, A. N. et Russell B. [1910], *14.02.

[82] Voir Russell [1903], Chap. XXXIII pour la définition des nombres réels à partir de segments de rationnels ainsi que Chap. XXXVI pour la continuité comme notion ordinale.

[83] On consultera Von Neumann et Morgenstern [1944], p. 630, note 1 pour une explication de l'utilisation de cet axiome dans un contexte comparable et des indications historiques à propos de son origine dans l'axiomatisation de la géométrie de Hilbert publiée en 1899.

[84] Voir Russell [1903], p. 254, pour une autre formulation de l'axiome d'Archimède. Nozick [1997], (1985) donne une formulation en termes de préférences. Comme elle utilise le concept de probabilité qui n'a pas encore été introduit, nous préférons, contrairement à Picavet [1996], p. 187, renoncer à cette interprétation dans le présent contexte.

[85] Voir Sobel [1990], p. 248 pour cette observation ou Fishburn [1970], p. 21 poùr la définition du concept de transformation affine positive.

[86] Ramsey [1926], p. 81, note 1.

[87] Le point de référence pour l'évaluation des gains et des pertes étant le moment du résultat du tirage, le coût du billet (1$) n'est pas considéré comme une perte à ce moment. Je ne débourse pas un dollar si je perds.

[88] Ramsey [1926], p. 82.

[89] Voir Ramsey [1926], p. 83 pour la preuve de (2) et p. 84 pour la preuve de (4).

[90] Une preuve de (5) se trouve dans Sahlin [1990], p. 38-39. Cette preuve est différente de celle que nous proposons. La transposée de (5) dans le langage des probabilités — ou Cr°

est remplacé par *Pr*, la disjonction est remplacée par l'addition de Hausdorff, et ainsi de suite — est discutée par Fishburn. Voir Fishburn [1980], p. 153.

[91] Plusieurs ouvrages d'introduction comportent une telle preuve. Ma version préférée est celle de Howson et Urbach [1989], p. 25.

[92] Voir Jeffrey [1965], p. 68-70.

[93] Ramsey [1926], p. 84.

[94] Voir Jeffrey [1965], p. 211, à propos du caractère trop « spacieux » du cadre bayésien.

[95] Pour une illustration de l'argument, voir Jeffrey [1965], p. 60-61.

[96] On se souviendra que nous avons ajouté (5), Ramsey ne mentionne que le respect des lois (1) à (4).

[97] Ramsey [1926], p. 84.

[98]Voir De Finetti [1935], p. 186 où le concept de cohérence est utilisé et défini. De même, De Finetti [1937], p. 103, ainsi que la note (b).

[99] Voir entre autres Jeffrey [1965], p. 51, Skyrms [1986], chap. VI, Earman [1992], p. 39, Resnik [1987], p. 75.

[100] On pourrait aussi suivre Picavet, Picavet [1996], p. 188, qui suggère l'expression « pompe à finance » pour traduire « *money pump* », un équivalent de « *dutch book* ».

[101] De Finetti [1937], p. 103.

[102] Pour la discussion des relations exprimées dans cette discussion sommaire du *dutch book*, on consultera surtout Skyrms [1986], p. 185 et Resnik [1987], p. 75.

[103] Voir Nozick [1993], chap. I et II ; la règle de calcul qu'il propose prend en compte la « valeur symbolique » d'une action et de ses conséquences. En mentionnant cette théorie à titre d'exemple, nous n'affirmons pas y souscrire sans réserves.

[104] Joyce [1999], p. 183. Isaac Levi ne serait pas d'accord. Voir Levi [2000].

[105] Voir Sobel [1998], p. 248.

[106] Ramsey [1926], p. 81.

[107] Sahlin et Gärdenfors [1988], p. 6 et 7, Jeffrey [1992]b, p. 215, Joyce [1999], p. 78. En conversation (janvier 2004), Joyce a reconnu que Ramsey [1926] ne comportait pas de théorème de représentation tout en affirmant qu'il est difficile de penser qu'il ne l'avait pas « rédigé et oublié dans un tiroir quelque part ».

[108] Sobel [1998], p. 251 ; Sobel évoque la possibilité qu'il faille élaborer et « peut-être augmenter » la liste des axiomes de Ramsey pour prouver un théorème de représentation qui soit bien spécifique à la théorie de Ramsey.

[109] Sobel [1998], et Fishburn [1981] discutent les axiomatisations voisines de celle de Ramsey dont celles proposées dans Pfanzagl [1967], et Debreu [1959].

[110] L'existence et l'unicité sont formulés ici d'après Sobel [1998], p. 248-249.

[111] Joyce [1999], p. 4.

[112] Voir les chapitres 8 et 9 de Jeffrey [1965] ainsi qu'aux autres références de la présente section.

[113] Il y a une différence dans le cas de ce que Savage appelle l'événement nul.

[114] Ramsey [1926], p. 79-80 ; *vide infra*, ce chapitre section 5 et *passim*.

[115] Davidson, D. et Suppes, P. [1956].

[116] Suppes [1984], p. 187.

[117] Sobel [1998] et Joyce [1999].

[118] Jeffrey [1965], p. 59 ; Jeffrey fait référence à ce que Sobel appelle les « quasi-mondes » *(near-worlds)* et les « mondes-valeurs » *(world-values)*.

[119] Von Neumann et Morgenstern [1944], p. 617-632.

[120] Le problème de cardinalité est mis en évidence, mais peu discuté, dans Fishburn [1981].

[121] *ibid.* p. 191.

[122] Mon intérêt pour les réels non-standards dans le contexte d'une théorie des préférences vient d'une explication donnée par W. Harper lors d'une conversation. John Bell a proposé l'analyse infinitésimale lisse pour exprimer la construction du continu à la manière d'Hermann Weyl dans une conférence intitulée « Weyl on intuition and the

Continuum » au colloque *Intuition in Mathematics and Physics* à l'Université Mc Gill (Montréal) en juin 1999. On consultera Bell [1998] et Bell [2003] pour une explication du concept d'analyse infinitésimale lisse. Bell [2000] montre l'application de cette approche au champ de la perception. Voir aussi la « théorie des quantités » de Brent Mundy, Mundy [1989].

[123] Un morceau d'information pourrait être un mot, un chiffre ou une syllabe. Voir Miller [1956] et Cherniak [1986] p. 47 *sq.*. Il existe des études plus récentes sur ce phénomène qui nuancent ce résultat. Voir Jonides [1995], p. 237 et *passim.*

[124] « While these theories may be satisfactory for normative purposes [...] » Davidson et Suppes [1956], p. 264.

[125] Ramsey [1926], p. 85.

[126] Jeffrey [1991], p. 29.

Notes du chapitre 3 :

[1] Une version antérieure de ce chapitre a été présentée au colloque sur la philosophie de R. Carnap tenu à Montréal en juin 1998.

[2] Voir « Intellectual Autobiography », Schilpp [1963], p. 74.

[3] Voir Carnap [1962]a.

[4] Il est néanmoins vrai que l'étude des mesures *a priori* de probabilité logique se prolonge dans quelques travaux de J. Hintikka, de T. A. F. Kuypers et E. T. Jaynes, travaux pour lesquels nous renvoyons le lecteur à la bibliographie de Howson et Urbach [1989] qui portent également un jugement peu enthousiaste sur cette orientation de recherche.

[5] Si on fait exception de la discussion de la théorie de l'objectiviste A. Wald dans un appendice, l'ouvrage qui traite du continuum des méthodes inductives, Carnap [1952], n'aborde pas la question du choix rationnel.

[6] Voir Carnap [1950], § 49.

[7] Carnap reprend probablement cette citation de Keynes, qui l'a rendue célèbre; elle vient de Joseph Butler [1736], *The Analogy of Religion.*

[8] Les observations empiriques (*evidence*) correspondent au paramètre *e* de $c(h,e)=q$, le degré de confirmation de l'hypothèse *h* sur la base des observations empiriques *e* est *q*.

[9] La formulation de cette fonction d'approximation est considérée comme un acquis de première importance dans le projet de construction de la logique inductive de Carnap [1950-4]. Cette fonction, dite « *c-mean estimate* », est la moyenne pondérée des valeurs possibles d'une grandeur où le facteur de pondération est le degré de confirmation. *op. cit.* § 99, 100A.

[10] Voir Carnap [1950-4], § 52, p. 280.

[11] Voir Carnap [1950], § 41 D.

[12] Dans la terminologie usuelle de la logique de la décision, c'est le principe de dominance faible.

[13] La loi de la diminution de l'utilité marginale est due à Daniel Bernoulli et Gabriel Cramer. Elle indique que "l'utilité en fonction de l'argent" est une fonction concave. Pour en savoir plus sur l'origine de cette observation si importante, on consultera l'historique de Todhunter [1865], p. 213 sq.

[14] Voir Carnap [1950], § 3, p. 7.

[15] Voir R. C. Jeffrey [1983], « Bayesianism with a human face », p. 101, chap. 5 de Jeffrey [1992]. Selon Jeffrey, il s'agit de l'essence même de la doctrine bayésienne.

[16] Carnap [1963]a, p. 74.

[17] Voir Carnap [1971], p. 7, note 1.

[18] Carnap [1950], § 12 « Psychologism in Inductive Logic ».

[19] Carnap [1962]a, p. 303.

[20] Carnap [1971], p. 14.

[21] Carnap [1960], p. 304.

[22] Dans Jeffrey [1965], p. 21, une distinction similaire est proposée entre ce qui est désirable et ce que l'agent désire.

[23] Comme Carnap l'indique dans Carnap [1971], p. 13, Bruno de Finetti a clarifié sa position en 1964, lors de la parution de la traduction anglaise de son article de 1937. Il y indique que la théorie des probabilités concerne le comportement cohérent et que le fait

que les gens soient eux-mêmes plus ou moins incohérents n'est pas essentiel. La première position de Bruno de Finetti se trouve dans De Finetti [1937].

[24] Carnap [1960], p. 309.

[25] Suivant M. Allais, on fait remonter cette idée centrale à l'essai de Daniel Bernoulli, Bernoulli [1738].

[26] Carnap [1971], p. 7.

[27] On en veut pour preuve les axiomatisations de Chernoff, Arrow, Hurwicz et Savage qui sont discutées au chapitre 13 de Luce et Raiffa [1957].

[28] Le terme est de Maurice Allais, un critique important de l'école dite « américaine ». Voir Allais [1979], version revue d'un article antérieur.

[29] Carnap [1963]b, p. 978.

[30] La bibliographie énumère plusieurs contributions de ces auteurs et d'autres associés à cette tradition. On hésite à mentionner Savage comme un « successeur » de Carnap.

[31] Voir Skyrms [1986], p. 206. Nous reviendrons ailleurs sur la représentation formelle de cette interprétation dite *d'extensibilité*.

[32] Voir von Neuman et Morgenstern [1944], § 3.2.1, p. 16. Nous exploitons différemment de ces auteurs l'analogie de la température que nous trouvons éclairante. L'exemple est également discuté par Carnap dans Carnap [1950], Chap.1 § 5.

[33] La fonction Cr est donc une mesure de probabilité, c.-à-d. une fonction de PROPOSITIONS dans NOMBRES RÉELS qui est non-négative et respecte l'additivité finie. Elle dite normalisée parce qu'elle assigne des valeurs entre 0 et 1, la valeur 1 étant celle de la proposition nécessaire.

[34] Voir Jeffrey [1983], p. 167. Jeffrey y discute une formule étonnante de C. I. Lewis, voir Lewis [1947] selon laquelle « pour qu'il y ait des choses probables, il faut quelque chose de certain ». Un *probabiliste radical*, tel Jeffrey, ne verrait aucun mérite à cette affirmation.

[35] Voir Carnap [1950], p. 306, les théorèmes T57-1b et T57-1d.

[36] Voir l'axiome A2 de Carnap [1971], p. 38, ou l'axiome A6 de Carnap [1963]b, p. 974.

[37] Carnap [1962]a, p. 308.

[38] Voir A. Shimony [1955].

[39] Cette argumentation bayésienne développe une suggestion proposée dans Jeffrey [1983].

[40] Pour une discussion informée, voir Howson et Urbach [1989], p. 99-105. Walliser et Zwirn [1997], p. 190-208, propose un répertoire varié de règles de révision.

[41] W. C. Salmon, Salmon [1988], p. 10.

[42] Voir, par exemple, Jeffrey [1983], p. 171 ou Salmon [1988], p. 10.

[43] Le terme grec *prohairesis* exprime le concept de choix délibéré, préférentiel, chez Aristote. Nous le préférons à l'expression « rationalité instrumentale » qui suggère une opposition entre les fins et les moyens qui est étrangère aux théories fondées sur le concept de préférence.

[44] On le constate, par exemple, en inspectant Carnap [1950-4] p. 343, Carnap [1955], Carnap et Jeffrey [1971].

[45] Voir Carnap [1955].

[46] Dans ce raisonnement, il est nécessaire de considérer que l'attribution des probabilités se fait en deux rondes successives.

[47] Voir Carnap [1950], § 62, p. 343.

[48] Ce sont les axiomes A7 à A10 de Carnap [1963]b.

[49] À ce sujet, il faut comparer le programme annoncé dans Carnap [1950] p. 562 et *sq.* pour le projet de ce qui devait être le Volume II de *Foundations* et l'introduction de Carnap [1952]. La difficulté est pleinement reconnue par Carnap ; voir Carnap [1971], p. 27.

[50] Voir Carnap [1963]b, p. 972.

[51] John G. Kemeny, dans Schillpp[1963], p. 711.

[52] Levi [1974], « On Indeterminate Probabilities », p. 129, chap 5 de Levi [1997]. Selon Levi, il y a des logiques inductives qui sont implicites dans les débats (sur les règles de révision), du côté des subjectivistes et du côté des objectivistes.

[53] Voir Carnap [1963]b, p. 973.

[54] Carnap [1963]b, p. 971.

[55] Carnap [1962]b, p. XV.

[56] Carnap [1963]b, p. 971-972.

Notes du chapitre 4 :

[1] Voir Fishburn [1988], p. 78.

[2] Parce que la date de première publication est 1954, nous utilisons le terme Savage [1954] bien que nous citions la seconde édition, revue et augmentée, parue en 1972. Il en va de même pour l'ouvrage de Jeffrey, Jeffrey [1965], pour lequel nous citons la 2e édition qui date de1982 avec les corrections de la réimpression de 1990.

[3] Voir Jeffrey [1965], p. 197-198 et [1992], p. 78.

[4] Ramsey [1926], de Finetti [1937], von Neumann et Morgenstern [1944].

[5] Savage [1954], p. 279.

[6] Luce et Raiffa [1957], p. 304.

[7] Savage [1954], p. 5.

[8] Savage [1954], p. 6.

[9] Savage [1954], p. 7.

[10] Savage utilise le terme « conséquences » (*consequences*) mais nous pensons, comme J. Joyce, que le terme «résultats», en anglais «*outcomes*» est préférable car il est plus neutre relativement à la connotation causale. Voir Joyce [1999] p. 48 note 3.

[11] Ces définitions sont, à quelques variantes stylistiques près, celles que l'on retrouve dans Savage [1954], p. 9.

[12] La notation que nous utilisons pour exprimer la théorie de Savage n'est pas exactement celle de Savage [1954]. Elle dérive plutôt de celle qui est utilisée par Luce et Raiffa [1957] et de Joyce [1999].

[13] Voir Joyce [1999], p. 83.

[14] Voir la critique de l'interprétation de Joyce dans Levi [2000]. Fishburn relève quatre interprétations différentes pour les actes de Savage ; voir Fishburn [1981] p. 143-144.

[15] « *A description of the world, leaving no relevant aspect undescribed* », Savage [1954], p. 9.

[16] Cette façon efficace et précise de présenter la trichotomie de Savage est due à Joyce, Joyce [1999] p. 49.

[17] Voir « Consequentialism and Sequential Choice », Levi [1997], p. 71.

[18] Suppes et Fishburn semblent voir les choses ainsi. Voir Suppes [1981] et Fishburn [1981].

[19] Voir Aumann et Anscombe [1963].

[20] C'est une des rares occasions où ne reprenons pas le terme proposé par Fishburn qui donne à cette propriété le nom de « Totalité »; Fishburn [1970] p. 24. Comparez à Jeffrey [1965], p. 145.

[21] Voir Savage [1954], p. 21.

[22] La formulation de cette définition se compare à l'explication donnée dans Savage [1954] p. 22. Voir la formule équivalente de Luce et Raiffa [1957], p. 302 ou Picavet [1996] p. 198.

[23] Luce et Raiffa [1957] p. 302.

[24] Savage [1954], p. 20 et 21.

[25] Voir Allais [1953].

[26] Les économistes qui formuleront des théories de l'utilité après Savage tels, par exemple Debreu [1960] prendront habituellement comme terme primitif la relation de préférence définie sur les conséquences. On doit cette observation à l'économiste Peter P. Wakker, 11e conférence on Foundations of Risk and Utility Theory, à Cachan, France (juillet 2004). On peut consulter Abdellaoui et Wakker [2004], p. 2 pour une définition de l'approche *outcome-oriented.*

[27] Voir le postulat SAV_5 de Joyce [1999] qui incorpore la définition D_s3.

[28] Luce et Raiffa [1957], p. 303.

[29] Voir Savage [1954], p. 31 et Fishburn [1981], p. 160.

[30] Savage [1954], p. 39, et pour l'inexistence d'actes d'utilité infinie, p. 81.

[31] Voir Savage [1954], p.77.

[32] On pourrait dire de façon équivalente, que u est unique relativement au choix arbitraire de l'unité et d'un point zéro.

[33] Voir Savage [1954], p. 34.

[34] L'idée vient de Jeffrey [1983], § 2, p. 82-88. Elle est transposée et commentée par Joyce [1999], p. 103-105.

[35] Joyce [1999], p. 83.

[36] Dans cette définition de A, on étend le langage de Savage ; pourrait suivre Joyce [1999] p. 64, et définir la conditionelle centrée $\ldots \Rightarrow \ldots$ comme une « conditionnelle AGM »; Gärdenfors [1988], p. 148 *sq.* Voir aussi Lindström et Rabinowicz [1989] et [1991].

[37] Joyce [1999], p. 67. Le meilleur candidat pour un tel acte serait la situation où quelqu'un forme l'intention de se suicider, *quoi qu'il advienne*. Même alors dit Joyce, il doit tenir compte de la possibilité que son suicide échoue. Ajoutons que cet argument, qu'on ne peut pas tout prévoir, est une caractéristique raisonnable de la délibération et que celle-ci est un argument contre le concept de choix absolument résolu.

[38] La monotonocité stricte, c.-à-d. « si $a \succ b$ alors $a(s) \succ b(s)$ pour tout état s »

[39] Abdellaoui et Wakker [2004] discutent des théories qui n'ont pas ce défaut et proposent leur analyse de l'incertitude centrée sur les résultats (*outcomes*).

[40] Shafer [1986], p. 206.

[41] Joyce [1999], p. 108.

[42] Picavet [1996] discute de l'échec de la transitivité des préférences dans le cadre de la *théorie du regret* de l'économiste R. Sugden. Picavet conclut en relativisant la valeur des postulats, trop souvent considérés (et testés) individuellement. *id.* p. 227-228. Voir aussi ce que dit D. Davidson au sujet des recherches empiriques conduites par lui-même et J. Marschak et en particulier, la difficulté d'interpréter les résultats en relation avec la validité des axiomes; Davidson [1980], p. 269 sq. Notre propre position a déjà été exprimée; elle est identique à celle de Jeffrey; voir Jeffrey [1992], p. 169, note 19.

[43] Cette version du problème est une généralisation de l'énigme proposée à l'origine par M. Allais, Allais [1969]. Les valeurs de l'exemple sont obtenues par une transformation linéaire des valeurs utilisées dans l'argumentation de Resnik ; voir Resnik [1987].

[44] Savage [1954], p. 101 et p. 102.

[45] Voir Maher [1993] à propos de situations qui concernent la transitivité de la relation de préférence.

[46] Savage [1954], p. 103.

[47] Ellsberg [1961].

[48] Resnik [1987], p. 105-107.

[49] voir Joyce [1999], p. 101-102., McLennen [1990], p. 70-71.

[50] Savage [1954], p. 14.

[51] Savage [1954], p. 82 et *sq.*, mais plus précisément p. 83.

[52] Savage [1954], p. 84.

[53] Picavet [1996], p. 197. Voir aussi la remarque sur la philosophie des passions chez Hume, *id.* p. 203. Cette citation va dans le sens de ce que nous disions au sujet de la *prohairesis*, chapitre 3, note 42.

[54] Jeffrey [1974], p. 226.

[55] Dans l'idiolecte de la théorie de la décision et de la théorie des jeux, l'expression « la nature » est utilisée pour désigner « le cours des choses », ce qui advient en dehors du contrôle de l'agent.

[56] Savage [1954] p. 23.

[57] Savage [1954] p. 83.

[58] Pour un survol des théories de l'utilité « dépendantes des états », du type de celle de Arrow-Debreu, on consultera Dreze et Rustichini [2004]. Voir aussi Wakker [2004].

[59] Voir Joyce [1999], p. 115-116 et Jeffrey [1965], p. 21.

[60] Contrairement à Joyce, nous formulons l'exemple de façon intemporelle pour être plus conforme à l'approche de Savage.

[61] Jeffrey [1965], p. 208.

[62] Voir Jeffrey [2004].

[63] Jeffrey fait l'historique de sa théorie dans Jeffrey [1965] p. xiii, et p. 149 ainsi que dans Jeffrey [1965]b et dans la préface de Jeffrey [2004]. Voir aussi Bolker [1966].

[64] En plus de son ouvrage majeur *The Logic of Decision*, (1965), 2e éd. 1983, ci-après Jeffrey [1965], Jeffrey a publié un recueil d'essais intitulé *Probability and the Art of Judgment*, ci-après, Jeffrey [1992] et de nombreux articles sur la logique de la décision, l'épistémologie bayésienne et les questions connexes.

[65] Jeffrey [2004].

[66] Pour les détails, autrement dit, l'essentiel, nous référons le lecteur aux définitions de Pollard [1999], de Jeffrey [1965], p. 148, de Bolker [1967], de Jeffrey [1974] et [1978] ainsi que les commentaires de Broome [1990], p. 484 et Joyce [1999], chap. 4, § 4.4, p. 127. Jeffrey [1978] p. 233 explique le sens exact de l'expression « relation représentable ». On remarquera en particulier la critique des restrictions présentes dans le théorème de Bolker et le problème de non-unicité noté dans Jeffrey [1974], p. 230 ainsi que dans Bolker [1967], p. 338 qui constitue une difficulté résolue subséquemment par Joyce pour sa propre théorie dans sa thèse de doctorat (1992) et dans Joyce [1999] en utilisant la suggestion proposée par Jeffrey.

[67] Jeffrey [1965], p. 82.

[68] Jeffrey [1965], p. 84. Selon Joyce, Jeffrey a renié cette idée plus tard. Joyce [1999], p. 149.

[69] Elle sous-tend le raisonnement « évidentiel » qui conduirait à choisir une seule boîte. Voir Joyce [1999], p.149 et sa critique de Jeffrey.

[70] Jeffrey [1965], p. 83 et 84.

[71] Voir plus haut, Chap. II, et pour un traitement différent de cette idée, ce que nous en dirons au chapitre VI ainsi que Vickers [2001].

[72] L'idée de valider un modèle théorique par des considérations de symétries est habituelle en physique comme le note van Fraassen [1986]. Il est raisonnable de penser que l'on pourra démontrer ce qui est affirmé ici au sujet des structures bayésiennes et nous tenterons de le faire dans une recherche ultérieure.

[73] Cette analogie m'est apparue très clairement en étudiant les critiques de la logique de la décision et la normativité de la logique de la décision, au début de ma recherche doctorale; d'où le projet de développer et de systématiser la logique de la délibération comme une véritable logique philosophique, une logique de la raison pratique. Nous avons découvert par la suite que la priorité de cette analogie revient à Jeffrey qui l'utilise à plusieurs endroits : Jeffrey [1965], p. 167 et chap. 12, ainsi que « Appendix » dans Jeffrey [1968] (p. 42 de Jeffrey [1992]).

[74] Jeffrey [1965], p. 211.

[75] Nous renvoyons le lecteur à Jeffrey [2004], aux travaux de Alan Háyek, Háyek [2001] et Háyek [2003] et aux travaux de Charles Morgan [2002] sur les probabilités conditionnelles comparatives.

[76] Jeffrey [2004], p. 13.

[77] David Lewis [1976], p. 137. Pour une démonstration courte et élégante, voir Jeffrey [2004], p. 15-16.

[78] On dit aussi « probabilistiquement indépendantes », un calque de l'anglais.

[79] Gärdenfors [1988], chap. 7, p. 147, en particulier le théorème 7.10, p. 158.

[80] Voir Joyce [1999], p. 22, théorème 6.2 pour un résultat de trivialisation analogue qui s'applique à ses connecteurs « $ » et « @ ». Joyce proposait un jugement similaire à celui que nous faisons ici dans sa conférence intitulée « Belief Revision Conditionals and Conditional Excluded Middle » dans le cadre de l'atelier « Frameworks for Inquiry : Representing Belief, Knowledge, and Conditionals », à l'université Western Ontario (London), en janvier 2004.

[81] On peut interpréter de cette façon, *grosso modo*, l'orientation de recherche de J. Joyce et de C. Morgan dans des travaux récents ou non encore publiés sur la logique des probabilités.

[82] Pour cette définition d'une option, voir Jeffrey [1974]b, p. 165 et comparer avec celle de Lewis qui est expliquée dans le même passage.

[83] Voir Halmos [1974] p. 5 pour les principales lois et l'axiomatisation d'une algèbre de Boole.

[84] Bolker [1967], p. 337.

[85] Bolker [1967], p. 336.

[86] Jeffrey [1965], p. 157.

[87] Nous rapportons les observations de Bolker, dans Bolker [1967]. Domotor [1978] contourne ce problème et définit les préférences sur des domaines booléens finis au prix d'une grande complexité.

[88] On peut contraster avec ce que dit Jeffrey au chapitre 11, p. 166.

[89] Jeffrey [1965], p. 209.

[90] Une algèbre de Lindenbaum est une algèbre de Boole $B_L = \langle L, \vee, \wedge, \neg, 0_L, 1_L \rangle$ définie sur un langage L, où le zéro, 0_L, est $(\varphi \wedge \neg \varphi)$ où l'unité, 1_L, est $(\varphi \vee \neg \varphi)$ et où les opérations s'interprètent comme des connecteurs propositionnels la conjonction $\wedge$, la disjonction $\vee$ et la négation $\neg$. Voir Chang et Keisler [1973], § 1.4.

[91] Voir Jeffrey [1965] p. 213 et Halmos [1974] p. 78-79 pour le théorème de Stone.

[92] Skyrms [1986]a, p. 59; c'est nous qui soulignons.

[93] Voir Jeffrey [2004] p. 104.

[94] Voir Yeghiayan [2000].

[95] Voir Jeffrey [2002] et Jeffrey [1992].

[96] Pour une argumentation soignée et convaincante contre cette solution au problème de Newcomb, nous renvoyons le lecteur à l'explication de Joyce; voir Joyce [1999], p. 154-161. Lewis a démontré — avec quelques bémols à la clé — que le dilemme du prisonnier est un problème qui, sur le plan formel est rigoureusement du même type que le problème de Newcomb; voir Lewis [1979]. Jeffrey croit aussi que certains dilemmes du prisonnier sont des problèmes de Newcomb; voir le problème d'Alma, Jeffrey [1988].

[97] Notre explication n'utilise que le principe de dominance. Pour un exemple d'analyse complète du dilemme avec une matrice de désirabilité au sens de Jeffrey avec une distribution de valeurs numériques qui assigne des probabilités aux actions de l'autre, voir Jeffrey [1965], p. 16-17.

[98] Jeffrey [1965], p. 19. Eells est du même avis : voir Eells [2000], p. 896.

[99] Voir Jeffrey [1965], (2^e édition), p. 20, l'exemple 14.

[100] Voir la démonstration dans Joyce [1999], p. 158-159.

Notes du chapitre 5 :

[1] Le problème de Newcomb tire son nom de celui qui l'a découvert, le physicien William A. Newcomb, chercheur au laboratoire Lawrence Livermore de l'Université de Californie. Pour l'historique, voir Nozick [1970].

[2] Le terme *prédicteur* est emprunté à la statistique et l'usage que nous en faisons ici est déviant ; ce terme désigne habituellement une variable plutôt qu'une personne.

[3] David Lewis signale, avec raison, que l'antériorité du moment de la prédiction n'est pas essentielle, pourvu qu'il y ait indépendance causale entre la prédiction et le choix de l'agent. Voir Lewis [1979], p. 300.

[4] L'idée de concevoir l'oracle comme un psychologue perspicace dont les conjectures sont hautement probables sans être certaines est due à Sobel. Voir Sobel [1988] pour la critique des prédicteurs infaillibles.

[5] Les expressions « *one-boxer* », « *two-boxer* » et « *no-boxer* » sont des idiomes pour lesquels il n'y a pas de traduction reconnue. Voir Picavet [1996], p. 244, note 1. Il nous semble qu'on perdrait une nuance importante en contournant le problème par une périphrase qui évite de qualifier l'agent lui-même.

[6] Selon Pearl [2000], plus personne ne soutiendrait aujourd'hui la théorie « évidentielle » telle que proposée par Jeffrey en 1965 et la solution one-boxer serait aujourd'hui universellement jugée indéfendable. David Lewis, Lewis [1981], fait un plaidoyer très solide contre l'insoutenable légèreté du principe MUE dans ce contexte.

[7] Nozick [1970], dans Moser [1990] p. 210.

[8] Le terme « évidentiel » est clairement un calque de l'anglais. Nous l'utilisons comme terme technique par opposition à « causal » avec la conviction ferme qu'il n'existe aucun équivalent français légitime, ni de périphrase dont le sens serait clair et l'usage établi.

[9] Selon David Lewis, Terence Horgan commet cette erreur. Voir Lewis [1981] note 18 et Horgan [1981].

[10] Joyce pose clairement cette clause comme *une donnée* du problème. Joyce [1999], p. 152.

[11] Ces remarques jouent le même rôle dans notre explication du problème de Newcomb que le *Good Sharp Observation Principle*, dans la « solution » de Keith Lehrer et Vann McGee. Voir Lehrer et McGee [1991].

[12] Voir Belnap, [2001], p. 182 (Post 3, § 7A.2) « *no branching toward the past* ». Lewis observe que la théorie soutenue par Terence Horgan dans Horgan [1981] ne respecte pas ce principe et comporte des conditionnelles contrefactuelles à effet rétrograde.

[13] Voir Nozick [1974], p. 81 et 82.

[14] Nous renvoyons le lecteur aux arguments de Belnap sur l'indéterminisme et « la fine ligne rouge », chapitre 6 de Belnap [2001]. Maitzen et Wilson rejettent le concept d'omniscience anticipative comme inintelligible et ils citent les articles de Patrick Grim qui est du même avis. Mentionnons aussi Sobel [1988] qui rejette l'idée de prédicteur infaillible.

[15] Le problème de Fisher est exposé et discuté dans le chapitre 1 de Jeffrey [1965]. On y trouvera la référence au document original de Fisher publié en 1959.

[16] Lewis [1981] p. 239 (§ 4).

[17] Voir Lewis [1981] p. 241 (note 8).

[18] Nous croyons que cette observation simple sur la répétition d'un problème de Newcomb indique une propriété importante du problème de Newcomb ainsi qu'une différence significative entre le dilemme du prisonnier et le problème de Newcomb. Elle vient d'une suggestion faite par David Gauthier lors d'une conférence donnée à l'UQÀM en 1996. Voir Gauthier [1996]. Une recherche attentive ne nous a pas permis de retrouver cette idée dans ses écrits— elle est pourtant dans l'esprit de Gauthier [1985]. Brian Skyrms (en conversation) nous a suggéré la possibilité qu'il s'agisse d'une idée de Douglas Hofstadter. Nous n'avons pas retrouvé cette idée dans les écrits d'Hofstadter non plus. Voici quelques références complémentaires : le concept de *super-rationalité* dans Hofstadter [1985], p. 748-749 ; la discussion des arguments de Kent Bach par Sobel, Sobel [1994], chap. 5 et la discussion de la position générale de Gauthier, dans Sobel [1994], chap. 6 ; le concept de *world decision problem* de Brian Skyrms, Skyrms [1982], p. 707.

[19] Voir Gauthier [1985], p.123 de Sobel [1990]a, (§ VI).

[20] Voir Skyrms [1980], p. 131.

[21] L'ambiguïté du problème de Newcomb a déjà été notée, voir Levi [1975]« Newcomb's many problems », Jeffrey [1965], p. 24. Plus récemment, Peter Schorer, dans Schorer [2003], aborde le problème de Newcomb comme un cas particulier de ce qu'il appelle les paradoxes de simulation. Voir aussi les arguments de Stephen Maitzen et Garnett Wilson, Garnett et Wilson [2003].

[22] Maitzen et Wilson [2003], p. 153.

[23] Voir Gupta [2000] ainsi que les références aux travaux antérieurs de Gupta qu'on y trouve.

[24] Gupta [2000] et Chapuis [2000] proposent des contributions pertinentes mais leur recherche sur la circularité du concept de rationalité conduit rapidement à considérer la théorie des jeux. En théorie des jeux, la connaissance commune est définie comme un ensemble de propositions contenant les règles du jeu ainsi qu'une description des faits pertinents pour le jeu pour lequel on construit l'ensemble des propositions P_n telles que le joueur 1 sait que le joueur 2 sait que le joueur 1 .... sait que P_n . Cette définition est manifestement circulaire, mais de façon différente de la circularité qui est présente dans le problème de Newcomb.

[25] Chapuis et Gupta [2000], l'ouvrage contient quelques articles qui partagent ce présupposé.

[26] Gibbard et Harper [1978].

[27] Lewis [1981]. Lewis expose les relations entre diverses formulations de la théorie causale de la décision, dont celles de Gibbard et Harper, Skyrms et Sobel. Dans leurs articles ultérieurs, Sobel et Joyce se situent relativement à la version de Lewis.

[28] C'est la formule de Gibbard et Harper mais la notation est modifiée. Voir Gibbard et Harper [1978], p. 352. Nous citons d'après la version publiée dans Gärdenfors et Sahlin [1988].

[29] Les expressions « conditionnelles contrefactuelles » et « contrefactuelles » sont des calques de l'anglais (*counterfactuals*). On constate que l'usage en est répandu en français.

[30] On pourrait soutenir que l'*imaging* est une façon de contourner la difficulté.

[31] Comme nous l'avons vu au chapitre IV, la conditionalisation bayésienne est la règle utilisée par Jeffrey $P(\mathrm{A/B}) = P(AB) / \mathrm{P}(B)$.

[32] Voir le théorème d'impossibilité de Lewis [1976] pour un conditionnel qui pourrait correspondre à la révision de croyance.

[33] Joyce signale que le tiers exclus conditionnel joue un rôle essentiel sur le plan formel; il est nécessaire pour que les conjonctions de subjonctifs qui représentent les options forment effectivement une partition. Voir Joyce [1999], p. 170.

[34] Lewis [1973], p. 80.

[35] idem. L'expression « Ce n'est pas le cas que » est inusitée en français. Ici, elle aide à bien indiquer la position de la négation dans la forme logique qui porte sur une proposition composée. Le bon usage voudrait garder la négation proche du verbe principal.

[36] Gibbard et Harper ont remarqué ces difficultés, op. cit. notes 2 et 3. Ils énoncent eux-mêmes les limites de la sémantique des conditionnelles contrefactuelles en signalant « la nécessité d'une approche plus générale ».

[37] Dans le contexte de la logique temporelle, l'idée selon laquelle, si le déterminisme est faux, il n'existe pas une totalité bien définie (un monde possible unique) qui *résulte* de l'acte accompli remonte à Prior [1956], p. 92 : « [...] there is no such totality » écrit-il. On retrouve cette même idée chez R. Thomason, N. Belnap, R. Stalnaker, J. F. Horty et D. Vanderveken. Contrairement aux apparences, ceci ne contredit pas l'idée « qu'il n'y a qu'un seul monde dans lequel nous vivons tous, le monde commun des processus physiques et de nos actions » Belnap, Perloff et Xu [2001], p. 177. C'est que les conséquences font partie du futur indéterminé.

[38] L'idée de *l'imaging* vient d'une suggestion de Skyrms et remonte à la fin des années 1970. Voir Lewis [1981], la section sur l'imaging et le concept d'imaging flou (*blurred imaging*) développé par Gärdenfors, Gärdenfors [1988], chap. 5. Voir aussi Sobel [1994], p. 152-165, Joyce [1999], p.172-176 de même que Lepage [2000], en particulier, la clause (5) et le paragraphe qui précède, page 37.

[39] Il y a plusieurs façons de réaliser cette caractéristique. Cette question est inséparable de la sémantique des conditionnelles ; elle fera l'objet d'une recherche ultérieure.

[40] Voir Lewis [1981], § 2, p. 237.

[41] On utilise le néologisme « conditionnaliser » en tant que terme technique, dérivé de l'adjectif et du nom « conditionnel ». En français, le terme « conditionner » possède un tout autre sens.

[42] Lewis [1981], § 2, p. 237.

[43] Ce sont les définitions de Lewis [1981], § 2.

[44] L'idée que la désirabilité de Jeffrey est invariante relativement aux partitions est expliquée informellement par Sobel dans Sobel [1986], p. 142, 161. Pour une définition formelle de l'invariance relativement aux partitions, voir Joyce [1999], p. 121, et surtout la discussion animée où Joyce critique Lewis et Sobel ; p. 176-177.

[45] Lewis [1981] expose de façon complète et détaillée cette analyse comparative.

[46] Voir la section 9 de Lewis [1981], p. 325 à 329 de Lewis [1986].

[47] Lewis [1981], p. 254.

[48] Lewis [1981], p. 246.

[49] Voir Lewis [1976].

[50] Joyce [1999], p. 150-151, et id. p. 116-117.

[51] Voir le classique Salmon [1970], puis Hitchcock [1993] et [1996] ainsi que Pearl [2000].

[52] On doit mentionner le théorème antérieur proposé par Armendt [1986] qui prend comme base la théorie de Fishburn. Voir Joyce [1999], p. 224-229 propose une critique.

[53] Joyce [1999], p. 20.

[54] Joyce [1999], p. 136. Voir aussi Bradley [2001], p. 283 pour une critique de cette thèse.

[55] Au sens de l'opérationnalisme que Percy Bridgman recommandait pour la physique : « un concept se réduit à l'opération qui permet de le définir ».

[56] Joyce [1999], p. 21. Voir aussi McCall [1999] qui établit l'importante distinction entre une raison de délibération et une raison de justification.

[57] Joyce [1999], p. 41.

[58] Joyce [1999], p. 138.

[59] Il existe plusieurs recensements de Joyce [1999]. Mentionnons Levi [2000], Eells [2000], Bradley [2001]b, Fitelson [2003] et Weirich [2000].

[60] Joyce qualifie la théorie de Jeffrey de « théorie de la valeur » et lui refuse le titre de « logique de la décision » dans le résumé d'une conférence pour un atelier sur l'épistémologie bayésienne au 26^{ième} symposium international sur Wittgenstein, Kirchberg, Austriche, Août 2003. Voir aussi Levi [2000], note 9.

[61] Jeffrey l'avait fait dans Jeffrey [1995]. Dans Jeffrey [2004], à la suite de Skyrms et de Jeffrey [1996], énonce cinq principes qui permettent d'analyser et d'expliquer le problème de Fisher.

[62] On dit que la décision de faire un acte A fait écran (*screens off*) à toute corrélation purement évidentielle entre les actes et les états du monde si pour tout acte B et pour tout événement E, $P(E \mid B \,\&\, \boldsymbol{d}(A)) = P(E \mid \boldsymbol{d}(A))$, où $\boldsymbol{d}(A)$ est la décision de faire A, et $P(E \mid B)$

la probabilité conditionnelle de E quand B. Dans sa délibération, l'agent a déjà pris en compte l'information (la *news value*) qu'il va faire l'acte. Voir Jeffrey [2004], chap. 6.

[63] Voir Joyce [1999], p. 4 à 7, p. 79, p. 119 et p. 161 ; aussi, Joyce [2002].

[64] Joyce [1999] comporte aussi une contribution originale à la théorie des conditionnelles en fournissant une caractérisation formelle et les bases d'une analyse des « fonctions de supposition » qui permettent de représenter et de contraster des règles de révision des croyances. Joyce [1999], p. 189-195 et Joyce [2004].

[65] Une σ-algèbre A définie sur un ensemble S est une famille de sous-ensembles qui est caractérisée par trois conditions : (1) Elle contient l'ensemble vide. (2) Pour tout ensemble $E \in A$, le complément de E est aussi élément de A. (3) A est fermé sous l'union dénombrable.

[66] Nous pensons en particulier à la recherche de Thomason et Horty [1996], ainsi qu'à d'autres travaux appartenant au programme de la théorie qualitative de la décision qui seront mentionnés au chapitre VI.

[67] Voir Joyce [1999], p. 195 – 200, 257. Dans son ouvrage, Joyce a suggéré qu'une certaine forme d'imaging, qu'il appellera plus tard « l'imaging de Laplace » pourrait être la bonne interprétation des conditionnelles pour représenter le raisonnement causal à propos des effets des actions. Il croit maintenant que cette approche ne tient pas la route. Joyce [2004], p. 10 : « This was an error ! ».

[68] Joyce, [1999] p.70-73, et p. 176.

[69] Selon Joyce, sa théorie permet à l'agent d'associer des probabilités à ses propres actions, elle ne le force pas.

[70] Ce slogan de Levi, voir Levi [2000] est devenu le nom de ce problème et il est repris dans Joyce [2002] Rabinowicz [2002] et Jeffrey [2004].

[71] Spohn [1977], p. 114. Ledwig [1998] fait la synthèse du point de vue de Spohn et le compare avec celui de Jeffrey. L'auteur observe et distingue des fluctuations sémantiques importantes à propos de l'expression « associer une probabilité à des actes ».

Notes du chapitre 6 :

[1] *Éthique à Nicomaque*, livre III, 5 1112a20, 1112b13.

[2] Lewis [1979]a, p. 149.

[3] Voir Pollock [1991].

[4] En exprimant ces réserves nous songeons à Yudkowky [2002], par exemple.

[5] Par exemple, Dastani et al. [2003].

[6] Dastani et al. [2003]. p. 23.

[7] Pour Richard Jeffrey, les références sont Jeffrey [1988], Jeffrey [1988]a, Jeffrey [1995], Jeffrey [1996] et Jeffrey [2004] chap. 6 ; pour Brian Skyrms, Skyrms[1986]a, Skyrms[1987], Skyrms[1988], Skyrms[1990] ; pour Edward McLennen, McLennen [1990].

[8] Nous discutons ailleurs du débat entre Joyce, Levi et Rabinowicz. Levi [2000], Joyce[2002], Rabinowicz [2002].

[9] On doit la formulation de ces contraintes à Brian Skyrms.

[10] Voir Jeffrey [1995].

[11] Jeffrey [1996], p. 9. La formule Pr (A | B & C) = Pr (A & B | C) / Pr (B | C) est l'égalité qu'il faut postuler pour déduire « l'écran » des deux autres contraintes.

[12] Nous devons remercier R. Jeffrey (en conversation en 1996) de nous avoir orienté vers le concept d'écran (*screening off*) à une époque où nous tentions d'articuler notre projet de recherche et de cerner l'essentiel dans l'énigme de Newcomb.

[13] Jeffrey [2004], p. 107.

[14] Il n'y a rien de philosophiquement « profond » dans cette manœuvre de réécriture. Elle découle de la possibilité de paraphraser subjectivement tout jugement de valeur qui a une prétention objectiviste. Le langage des « états profonds » a la même capacité expressive.

[15] Jeffrey [1993], p. 143-144.

[16] Le « principe principal » de D. Lewis qui permet de définir le concept de « chance objective » à partir de la probabilité subjective semblait indiquer un pas dans cette direction. Voir Lewis [1980]

[17] Jeffrey [1995], p.7.

[18] En anglais, on dirait « *the base logic* ».

[19] La possibilité d'utiliser les *fonctions de choix futurs* pour exprimer l'indépendance causale et relier la logique modale et la théorie de la décision avait été évoquée par Thomason [1984] à la suite de Thomason et Gupta [1980].

[20] Prior [1967] et Thomason [1970] et Thomason [1984].

[21] Les collaborateurs mentionnés dans Belnap, Perloff et Xu [2001] sont assez nombreux ; P. Bartha, M. Green, J. Horty sont des co-auteurs.

[22] Belnap, Perloff et Xu [2001].

[23] Nous croyons que cet ouvrage de référence servira de mode d'emploi (*shop manual)* pour construire plusieurs autres logiques intéressantes dans les années à venir.

[24] Comme le note Belnap, ce sont les histoires qui se ramifient et non le temps lui-même. Belnap et al. [2001], p. 29 note 1.

[25] Le lecteur qui est déjà familier avec la logique du temps ramifié et la logique de l'action avec l'opérateur *stit* trouvera peu d'information nouvelle dans le reste de cette section.

[26] Nous avons été initié aux rudiments de la logique du temps ramifié dans le séminaire de logique philosophique de Storrs McCall à l'université McGill.

[27] Horty [2001], p. 6.

[28] Ou l'équivalent notationnel, Horty [2001].

[29] Il s'agit d'éviter le sophisme « *post hoc ergo propter hoc* ». Belnap, Perloff et Xu [2001], p.180.

[30] L'expression « attrition des branches », (*branch attrition*) est utilisée par S. McCall dans McCall [1994].

[31] McCall [1990].

[32] Horty [2001], p. 12.

[33] Cette réserve est exprimée, de la même façon dans les deux ouvrages. Belnap et al [2001], p. 33 ; Horty [2001], p.12.

[34] Belnap et al. [2001], p. 211.

[35] von Neumann et Morgenstern [1944].

[36] Voir Belnap et al [2001], p. 342.

[37] Jon Barwise exploite bien l'image du plan d'une ville pour illustrer la relation entre l'information et le monde. Voir Barwise, J. et Seligman, J. [1997] : *Information Flow: The Logic of Distributed Systems.*

[38] MacCall est particulièrement explicite sur la structure de continuité du temps linéaire des histoires. Voir McCall [1990] ou McCall [1994].

[39] On trouve une discussion de cette question dans Belnap et al. [2001]. Voir aussi von Neumann et Morgenstern [1944], p. 77-78 ainsi que Shenoy [1988].

[40] Cette façon inhabituelle de décrire une partition est celle que von Neumann et Morgenstern utilisent à propos des arbres. Voir von Neumann et Morgenstern [1944] p. 67.

[41] McCall [1994] exploite ce fait pour proposer une nouvelle interprétation des probabilités objectives fondées sur la proportionnalité des branches.

[42] Thomason indiquait simultanément la possibilité de clarifier la notion de dépendance causale à l'aide du modèle et le fait que cette clarification n'avait pas encore été réalisée. Thomason [1984], p. 159.

[43] Voir Velleman [1993] et Joyce [2002] ; l'idée de l'autorité épistémique de l'agent à propos de ses états intentionnels est déjà présente dans Anscombe [1957].

[44] Dans ce contexte le fait d'utiliser l'expression « acte mental » n'équivaut pas à l'expression d'une prise de position dualiste. Il se peut que la neurophysiologie, avec ses scanners FMRI nous permette bientôt de situer les processus nerveux qui correspondent aux décisions dans le cerveau. De façon étonnante, on peut déjà visualiser l'excitation des centres neuronaux qui correspondent aux préférences d'un sujet.

[45] Cette continuité entre les actes mentaux, telles les décisions, et les actes physiques, tel lever le bras, nous semble une caractéristique importante que nous devons à Daniel Vanderveken. Elle permet de penser l'engendrement des actes dans une optique de continuité.

[46] I. Levi et W. Spohn ne seraient pas d'accord avec nous. Voir Joyce [2002].

[47] Joyce [1999], p. 156 : « Soit **d** *A*, un terme qui dénote la décision de faire *A* ».